

STIC-Biotech/ChemLib

72221

From: Slobodyansky, Elizabeth
Sent: Wednesday, July 31, 2002 11:44 AM
To: STIC-Biotech/ChemLib
Subject: 09/900,038

Please search for case 09/900,038:

SEQ ID NOs: 1 and 2 against commercial and ^{ES} ~~interference~~ databases.

Thank you.

Elizabeth Slobodyansky, PhD

Primary Examiner

Art Unit 1652
CM1 10D11
703-306-3222

mail box 10C01

CRFE

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: _____
Date Completed: 8/6/02
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

Point of Contact
P Sheppard

Phone number: (703) 308-4499

TYPE OF SEARCH:

NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)

STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 2, 2002, 10:43:39 ; Search time 1790.74 Seconds
(without alignments)
10973.127 Million cell updates/sec

Title: US-09-900-038A-2

Perfect score: 939

Sequence: 1 atgaattatgattcattat.....gaggagagaacaaagtgc 939

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_hgt.*

3: gb_in.*

4: gb_cm.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pi.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pi.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vi.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htg_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
------------	-------	-------------	--------	----	-------------

ALIGNMENTS

RESULT 1

AB050723

LOCUS

DEFINITION

AB050723

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

AB050723 9987 bp DNA linear BCT 03-FEB-2001
Streptococcus agalactiae cpsIBD, cpsIBF, cpsIBG, cpsIBH, cpsIBI,
cpsIBJ, cpsIBK, cpsIBL, neuB, neuC genes, complete cds.

AB050723.1 GI:12697593

Streptococcus agalactiae DNA.

Streptococcus agalactiae

Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;

Streptococcus.

1 (sites)

Miyake, K., Watanabe, M. and Iijima, S.

CpsJ of Streptococcus agalactiae type Ib shows

beta-1,3-galactosyltransferase activity

Unpublished

2 (bases 1 to 9987)

Miyake, K., Watanabe, M. and Iijima, S.

Direct Submission

Submitted (01-NOV-2000) Katsuhide Miyake, Nagoya University, Grad.

Sch. of Engineering, Dept. of Biotechnology; Chikusa-ku, Furo-cho,

Nagoya, Aichi 464-8603, Japan

(E-mail:miyake@proc.nubio.nagoya-u.ac.jp, Tel:81-52-789-4278,
Fax:81-52-789-3221)

FEATURES	source	Location/Qualifiers
	RBS	1..9987 /organism="Streptococcus agalactiae" /db_xref="taxon:1311"
	gene	425..430
	gene	442..1791
	CDS	/gene="cpsIbD" 442..1791 /gene="cpsIbD" /function="glucosyltransferase" /codon_start=1 /transl_table=11 /evidence=not_experimental /product="CpsIbD" /protein_id="BAB21597.1" /db_xref="GI:12697594"
	RBS	YRDMSRGYLEEFKMLVRSFYIIPISLSLFFISKNSFTTTLRSLPFTFIAMNSILLYL LNSFLKYRYKYSYAKFSRDTKVLITNKSLSKMTFRNKYDHNYYIAVCILDSSEKDCY DLKHSNLRINKDALTSBTLCTVDOAFINIPIELFGYQIQDIINDIEAMGVIVNVN VEALSFNIGEKRIQTFEGYSVITYSMKEYKSHLIAKREFLDIMGAILGLLIGCVAI FLVPIQRDGGPAIFSQNRVGRNGRIFRKYFRSMRVDAEQKKOLLVHNMQTMFMFK LDDDPRIITKIGKFIKRTSIDELPQFINVLKGDMSLVGTRPPTVDEYKYNSTQKRRLS FKPGITGWIQISGRNNITDFDEIVKLDVQYINNEWSIWSIDIKIILITLKVVLGTGAK"
	gene	1815..2264
	gene	/gene="cpsIbF" 1815..2264 /gene="cpsIbF" /codon_start=1 /transl_table=11 /evidence=not_experimental /product="CpsIbF" /protein_id="BAB21598.1" /db_xref="GI:12697595"
	CDS	/translation="MKICLVSSGGHLAHLNLLKPIWEKDRFWVTFOKEDARSLRE ETVHCFPTNKNVKNVKNITLAFKVLKRKRPDVIISGAAVAPFFVIGKLFCKT VYIEVFDRIKPTLGLVYPVTDKFIQVQWEEKVKYPKAINLGIGF" 2249..2256 /gene="cpsIbF" 2264..2746 /gene="cpsIbG" 2264..2746 /gene="cpsIbG" /function="beta-1,4-galactosyltransferase" /codon_start=1 /transl_table=11 /evidence=not_experimental /product="CpsIbG" /protein_id="BAB21599.1" /db_xref="GI:12697596"
	RBS	/translation="MIFVTVGTHEQQFNRLIKEVDRLKGTGAIDQEVFIQTGYSDFEP QNCQWSKFLSYDDMSYMKAEIVITHGGPATFMNVAISKGRXTIVVPQEPQGEHVN HOVDLKEFLFYELDYILNITSELENIIEKNISTSKVISQNNDFCSSFKNELSKLFE "
	gene	2806..2812
	gene	2842..3981
	CDS	/gene="cpsIbH" 2842..3981 /gene="cpsIbH" /function="putative CP polymerase" /codon_start=1 /transl_table=11 /evidence=not_experimental /product="CpsIbH" /protein_id="BAB21600.1" /db_xref="GI:12697597"
	RBS	/translation="MQLLLALIVLLICSSYNEKMKFLNMAEIFFIVFMVYLVSI LNSLFRSEFHFVIAFNSLAVGVSVLLFYHYKNTNIELKLLKSLFNATILFCIG FLYYAIYFDVENVSLFGRNLIGSDWINGMHTORAMAFEYSNLLIPLITIINIYII YIKQYISSGMILGALLSTIILPIGGSGRAGIIVVLVQLVILLNNTIIVIKROTIRFEL
	RBS	YLVPTILLVILREDNLVSIYNRINIIRSGSSSRFSLYKDYVHSVITDLSFLGKGY RELWNSDLPLGSHSTYIGFYKTLGFLGINVLGLFLLILISIIKEAKKSDFYIEIVG SVILLFSFALEDIDGANWLIIFVFTVLGILENKDFYSQLKRWES"
	gene	3964..3970 /gene="cpsIbH" 3981..4955 /gene="cpsIbI" 3981..4955 /gene="cpsIbI" /function="beta-1,3-N-acetylglucosaminyltransferase" /codon_start=1 /transl_table=11 /evidence=not_experimental /product="CpsIbI" /protein_id="BAB21601.1" /db_xref="GI:12697598"
	gene	/translation="MEKQILVSIPIYVNSEAYLKECVOSVLOQTHSLSEVILINDGS TDNSGEIDNLSQKDDRILVFKHKNKGVSARNLGDKSTGEFITFVDDDFVAPNII EIMLNLTEDADIAEVDIFDIERNDRYKRRNFYKVFKNNSLKFEISGRVENVIV CTKLKKSITGNLREDENLKIGEDLLFNCKILCQEHCIWVDTTSSLYTVRIYKTSAMN QEFNSLDQFITFENEISIVPAKLANVVEAKFLREKVKCLRMKMFELGNSIDSKIKIQ REIFFKDKVLPFYKAVKYLKGLLSIYLMKCSPLIYIKLYDRFQKQ"
	RBS	4995..5001
	gene	5008..5949 /gene="cpsIbJ" 5008..5949 /gene="cpsIbJ" /function="beta-1,3-galactosyltransferase" /codon_start=1 /transl_table=11 /evidence=experimental /product="CpsIbJ" /protein_id="BAB21602.1" /db_xref="GI:12697599"
	CDS	/translation="MNYSIIMSVYNEPLNYVRDSVESILNQILTDFEFTIIVDNPISRG DLKQFLTYSVVDNRKIKILLNEENIGLASSLNKAVKISGGEYIFRMDADDISYPSRFD KOIRMEINSLDFSATLIELIDQKGNLVYKQESNKIYIINDIRMLNRSILAHTPW CVKVKVFKLMGYRDLVPVEDYDFAIRGALADFKIGLLNKVLLQVLRNENGISQTNKF KQYIYSAILODFYKESYIDITKITYFOEYVIKKRYTOQLSKYFELKSTPSITIRK LYICLYLVEKSPLVRRLLINDINILVCLKFGGEKQSD" 5928..5935 /gene="cpsIbJ" 5942..6880 /gene="cpsIbK" 5942..6880 /gene="cpsIbK" /function="putative sialyltransferase" /codon_start=1 /transl_table=11 /evidence=not_experimental /product="CpsIbK" /protein_id="BAB21603.1" /db_xref="GI:12697600"
	RBS	/translation="MTNRKIYVCHTYLHLLICLYKEEYISNLEIILSSSIDVDNLEK KUKSTINIHLEESGESELLSVLKDAGLSYKPSNCFIFNDATPIGRLLIKRHGI LYNLLEDGNCFYIFSQKLMKYVKKYILHKIQPHGFSRYCLGIEVNSLIDKPSDP RYKKEFVPRKELFDNVEYQKEMAINLFGAVRVSITKSPSVLVLQPLSDKEFMSYN NKIETSEOFNFYKSVINEYINKGVNVLKVPDRVDYDYSKLPVELLPSNVPMIEIEL MLTGREGGITHSSTALDFTCVDKKITVLVDLKDIK"
	gene	6865..6870
	gene	6877..8277
	CDS	6877..8277 /gene="cpsIbL" /gene="cpsIbL" /function="putative repeat unit transporter" /codon_start=1 /transl_table=11 /evidence=not_experimental /product="CpsIbL" /protein_id="BAB21604.1" /db_xref="GI:12697601"
	RBS	/translation="MNTKKLQSGFIYITGNLLVOGLAFITLPIYTRVISTEYVGQYS


```
LYVAMNIIIMFLICLOTSGSSARVKYGEERKSYSGSAFSGVNIWFLVLLIAFLFR
SLFAPLWFSFESIFLLMVCOSYANYVTFQCYFIOORSLANLILSLANAVSSVALS
LFLEFHSDDFLSVLGVFVPTIITGIVAFAYIYHSKSFYNPKYPRFIVTVSPLIF
HLGHOLLNOLDRLIMLARLYSTKEVAMYSFGLMHIQIIVLANSINMANWIPWEDARK
EKQLQSTYSRLYLGLVFLTGLTFVPELAQIMGGDQYSSSVQFISLIVSFLVF
LYTFPNNIOFYFANTWNPITGLLAAGVNNLNLVLPHIYAAAGAAATIIISYLALI
FHHVSKRVYHSDVSVRQVIILSGIVFSYAMLNMFNDINIVRWSLGIILIVYSIV
FORVILDLLSKRRRR"
8262..8269
RBS /gene="cpsIBL"
8277..9302
gene /gene="neuB"
8277..9302
CDS /gene="neuB"
/function="sialic acid synthetase"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="NeuB"
/protein_id="BAB21605.1"
/db_xref="GI:12697602"
/translators="MYIIAIEICNHNGDINLAKKVVAVSCGVDAVKFQTPKAEKL
ISFPAKAEYKATGTGADSOLEMKRLELSFEYLEMRYAISKGVETFTSPDEES
LEFLISTDMPYIIPGGEITNLPYLEIKGQKKVILSTGMAYNEEHQAVNLRQNG
TTDISILHCTTEYTPYPSLNLNVIHTLKDFDKDLTIGYSDHSIGSEVPFAAAMGAE
VIBKHTLDTNMEVPOHKASATPDILAAVLKGFALLNQALGRFEKIPDPVEEKNKIVA
RKSVALKPIKKGDIYSIENITVKRPGNGISPMWYDILGOEAQDDFEDEVIRDSF
ENQLPEL"
9367..9374
RBS /gene="neuB"
9379..9987
gene /gene="neuB"
9379..9987
CDS /gene="neuB"
/translation="100.0%; Score 939; DB 1; Length 9987;
Query Match 100.0%; Best Local Similarity 100.0%; Pred. NO. 2.5e-134;
Matches 939; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 atgaattatagatcattatgcgtatataatagcgtcttaaatattatgagagattca 60
Db 5008 ATGAATTATAGTATCATATGTCGGTATATATGATGAGCCTTTAAATATATGAGAGATTCA 5067
Qy 61 gtagaattatattaaatcaaacgcttactgattttgagttcataaattgtcattgataat 120
Db 5068 GTAGAATCTATATTAATCAAAACGCTTACTGATTTGGTTGCATATATGTCATTGATPAA 5127
Qy 121 ccaagtagaggtgatttaaaagcaattcttaacagaaattcagttgtagataaagaata 180
Db 5128 CCAAGTAGAGGTGATTTAAAGCAATTTCTTAACAGAATATTCAGTTGTAGATAAATAGAATA 5187
Qy 181 aaatcttgcttaagaataatattggttttagcatcaagatttgaacaaagcggtgaaa 240
Db 5188 AAAATCTTGCTTATGAAGAAAAATATGTTTACATCAAGTTTGAACAAACCGGTGANA 5247
Qy 241 atttcaaggagagaataatttttagaattgagtgatgatttcatttcataatccaagttaga 300
Db 5248 ATTTCTAAGGGAGAGATATATTTTAGAATGGATGCTGATGATATTTTCATATCCAAGTAGA 5307
Qy 301 tttgataagcaaaatcgttttatgagagaaatcattggtattctcagcaactctaata 360
Db 5308 TTTGATAAGCAAAATTCGTTTTATGGAGGAAAATTCATTGGATTTCTTCAGCAACTTAATA 5367
Qy 361 gaattgatagaccgaaagaaatttagtatataaacaacgagaaagataataaataatcac 420
Db 5368 GAATTGATAGACCAAAAGGAAATTTAGTATATAAACACGAGAAAGTAATAAATATATAC 5427
Qy 421 ttaactaatgatatacaggaagtgttattgaatagatctataacttgcaccaccaactgg 480
Db 5428 TTAACATATGATATACGGAAGATGTTATTGAATAGATCTATACITGCCACCCACCGTGG 5487
Qy 481 tgcgtaaaaaagaagtttttcgataaagttaattgagatatagagatttagtacctgttgaa 540
Db 5488 TGCGTAAAAAAGAAAGTTTTCGATAAGTTTAATGGGATATAGAGATTTTAGTACCTGTTGAA 5547
```

```
Qy 541 gattatgattttgcaataagagagctctgcgtattttcaaaatcgcttactcaataaa 600
Db 5548 GATTATGATTTTGCATTAAGAGGAGCTCTGGCTGATTTTCAAAATCGGCTTACTCAATAAA 5607
Qy 601 gtactttcacagtagattataaacgagagatggaatcacacaacccaataaagtttaagcaa 660
Db 5608 GTACTTTTACAGTATAGATTAAACGAGAGATGGAATATCACAAACCAATAAGTTTAAGCAA 5667
Qy 661 tatatttactcagctatttttcaagagattttataaagaaaaaatcttatatgatatacaca 720
Db 5668 TATATTACTCAGCTATTTTACAAGATTTTATAAAGAAAAATCTTATATTGATATCACA 5727
Qy 721 aaaattactaattactttccaagagatgctgataaagaacgctatactacagcaagctc 780
Db 5728 AAAATTACTAATTACTTTCAAGAGTATGTGATAAAGAANAACGCTATACACTCAGCAAGCTC 5787
Qy 781 tctaaatattttgagctaaaaatctacccttagtattactattagaaaaactatataattgt 840
Db 5788 TCTAAATATTTTGAGCTAAAATCTACCCCTAGTATTACTATTAGAAAACTATATATTGTT 5847
Qy 841 ttataatttatacttttaagctcccttggttaggaggttattataaataaataatattatt 900
Db 5848 TTATATTTTATACTTTAAAGTCTCCCTTGCTTAGGAGGTTATTAAATAAATGATATTATATT 5907
Qy 901 ttagtactgaaattggttggagagagaaacaaaagtgcac 939
Db 5908 TTAGTACTGAAATTCGTTTGGAGGAGAGAAACAAAGTGAC 5946
RESULT 2
AF337958 16448 bp DNA linear BCT 27-FEB-2001
LOCUS Streptococcus agalactiae strain NT6 type VI capsular polysaccharide
DEFINITION synthesis operon, complete sequence.
ACCESSION AF337958
VERSION AF337958.1 GI:13022158
KEYWORDS
SOURCE Streptococcus agalactiae.
ORGANISM Streptococcus agalactiae.
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
REFERENCE 1 (bases 1 to 16448)
AUTHORS McKinnon,K., Chaffin,D.O. and Rubens,C.E.
TITLE Direct Submission
JOURNAL Submitted (19-JAN-2001) Infectious Disease, Immunology and
Rheumatology, Children's Hospital and Regional Medical Center, 4800
Sand Point Way NE, Seattle, WA 98105, USA
FEATURES
source
1..16448
/organism="Streptococcus agalactiae"
/strain="NT6"
/serotype="VI"
/db_xref="taxon:1311"
895..16444
/gene="type VI capsular polysaccharide synthesis operon"
895..2425
/gene="cpsA"
895..900
/gene="cpsA"
919..924
/gene="cpsA"
956..961
/gene="cpsA"
968..2425
/gene="cpsA"
/feature="similar to LytR"
/codon_start=1
/transl_table=11
/product="CpsA"
/protein_id="AAK11658.1"
/db_xref="GI:13022159"
/translation="MSNHSRRHKKHSHTPLRVINLFLVIFILLVSVVSLFLMYRHHF
```


/codon_start=1
/transl_table=11
/product="beta-1,3-galactosyltransferase CpsVIJ"
/protein_id="AAK11667.1"
/db_xref="GI:13022168"
/translation="MSIYNELPKYIKSSIDSLISQILKNFEFIIIVNDPREDVYDILQTYLSDARVKVLSNERNGLAYSLNKAIKYSGNVIYFRMDADDISYPERFQIINFM EKYRLDSATLIEVMDEDEETLYKORINKNYFESDIYRLMSNRCVLSHPWCVKRAV FLRLDGYRNVPAEDYDFLLRLALSGIKCLIGLIDLLRYSNANGISQLININSIVTQ LYPRIKSKMDKC"
RBS 9965..9971
gene /gene="cpsK"
9965..10913

Query Match	38.7%;	Score 363.4;	DB 1;	Length 16448;	
Best Local	Similarity 62.4%;	Pred. No. 8.5e-47;			
Matches 585;	Conservative 0;	Mismatches 351;	Indels 1;	Gaps 1;	
Qy 2	tgaattatagatcattatgcggtataatgagcctttaaatattgtgagagattcag	61			
Db 9043	TGAAGTATAGTGCATCATGTCAATTTATAATGAACCATTTGAAGTACATTAATCTTCAA	9102			
Qy 62	tagaattatataaatacaacgcttactgatttttgagtttcataattgtcattgataac	121			
Db 9103	TAGATTCAATTTTATCAACAACATPAAAAAATTTGAATTTATTTATTTGTGATTGATAACC	9162			
Qy 122	caagtagagtgatttaagaacaattcttaacagagaatttcagttgttagataagaataa	181			
Db 9163	CTGCACAGAGATGCTATATGCATATTTTGCNAACATATCTAGCAAGCGATGCAAGAGTAA	9222			
Qy 182	aaatcttgcttaagaagaataattggtttgagcatcaagttttgaaacaaagcgtgaaaa	241			
Db 9223	AAGTATTATCAATGAGAGAAATATGGGATTAGCTTAGTTTAAACAAGACTATAAAT	9282			
Qy 242	tttctaaggagaaatatatttttagaattgagtgatgctgatgatatatttcataccaagttag	301			
Db 9283	ATTTCAGTTGGAAATATATTTTGAATGGATGGCGGACGATATTTTCATATCCAGAAAGAT	9342			
Qy 302	ttgataagcaaatcgttttatgaggaataattcattggatttctcagcaactctaataag	361			
Db 9343	TTGAAAGCAAATTAACATTTATGGAAGATATCGTTTATAGACTTCTCAGCACTTTAATAG	9402			
Qy 362	aattgatagaccnaaaaggaatttagtatataaacaacgagaaagtaaaaaataact	421			
Db 9403	AAGTTATGGATGAAGATGAAGAAACATATATAAACAGAAATAATAAATAAAATATT	9462			
Qy 422	taactaatgatatacggagaagatgttattgaatagatctatacttggccaccaccaacgttgt	481			
Db 9463	TTTCAGAGATATTTATAGAATGCTCTCTAATAGATGTGTCCTTTACATCCCACTTGGT	9522			
Qy 482	gcgcataaagaagattttcogataagttaatgggatagagatttagtacctgttgaag	541			
Db 9523	GTGTTAAAGAGCAGTATTTTATAGGCTAGATGGATATAGAAACATTTGTTCTCGCAGAG	9582			
Qy 542	attatgattttgcaataagaggagctgctgctgatttcataaatcggttactcaataaag	601			
Db 9583	ATTATGATTTTACTAAGGTAGCGTGTGTCAGGAATTAATGTGGTTTAAATCGTGATA	9642			
Qy 602	tactttcacagtatagattaaacgagaatggaatatcacaaaccaataagtttaagcaat	661			
Db 9643	TATTGCTTAGATAGGAGTAATGCTTAATGGTATTTCAC-AACTAATCAATATAAACAGT	9701			
Qy 662	atattactcagctattttcaagaattttataaagaataatccttatattgatatcaca	721			
Db 9702	ATATACGCAGCTATATCTCAGATTTGGTTCAAAAATGGCAAAATGTTAATGTAGACT	9761			
Qy 722	aaattactaattactttcaagagtatgtgataaagaacgcgtactcagcaagagctct	781			
Db 9762	ATATTCTTATTTATCTTGAAAGCTATATAAATAGCAATAAATTTGTCAAAAACACTTAA	9821			
Qy 782	ctaaatatatttgagctaaataatcaccctcagtagtattactatttagaaaaactatatattgtt	841			
Db 9822	AAAAATACCTAGACTTTTAAATTTTATAGTAGTAGTCTCTATGCTTAAAAAAAGTCTTTTGGGAG	9881			

Qy 842	tatatattatactttaagtcctccctggtaggaggttattataataataatgattataattt	901			
Db 9882	GATTTTATTTTCAATATATCTTTTGCACGTCGATTAAATTAATCAAGATTTTATATATT	9941			
Qy 902	tagtactagaaaattgttgaggagagaaacaaagtga	938			
Db 9942	ACAAAATAAAATATTGAACGTAAGAGGACACAGA	9978			
RESULT 3					
LOCUS	U67549	14561 bp	DNA	linear	BCT 28-JAN-1998
DEFINITION	Methanococcus jannaschii section 91 of 150 of the complete genome.				
ACCESSION	U67549	L77117			
VERSION	U67549.1	GI:2826363			
KEYWORDS					
SOURCE	Methanococcus jannaschii.				
ORGANISM	Archaea: Euryarchaeota; Methanococci; Methanococcales;				
	Methanococcaceae; Methanococcus.				
REFERENCE	1 (bases 1 to 14561)				
AUTHORS	Bult,C.J., White,O., Olsen,G.J., Zhou,L., Fleischmann,R.D., Sutton,G.G., Blake,J.A., FitzGerald,L.M., Clayton,R.A., Gocayne,J.D., Kerlavage,A.R., Dougherty,B.A., Tomb,J., Adams,M.D., Reich,C.I., Overbeek,R., Kirkness,E.F., Weinstock,K.G., Merrick,J.M., Glodek,A., Scott,J.D., Geoghagen,N.S., Weidman,J.F., Fuhrmann,J.L., Nguyen,D.T., Utterback,T., Kelley,J.M., Peterson,J.D., Sadow,P.W., Hanna,M.C., Cotton,M.D., Hurst,M.A., Roberts,K.M., Kaine,B., Borodovsky,M., Klenk,H.P., Fraser,C.M., Smith,H.O., Woese,C.R. and Venter,J.C.				
TITLE	Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii				
JOURNAL	Science 273 (5278), 1058-1073 (1996)				
MEDLINE	96337999				
REFERENCE	2 (bases 1 to 14561)				
AUTHORS	Bult,C.J., White,O., Olsen,G.J., Zhou,L., Fleischmann,R.D., Sutton,G.G., Blake,J.A., FitzGerald,L.M., Clayton,R.A., Gocayne,J.D., Kerlavage,A.R., Dougherty,B.A., Tomb,J.-F., Adams,M.D., Reich,C.I., Overbeek,R., Kirkness,E.F., Weinstock,K.G., Merrick,J.M., Glodek,A., Scott,J.D., Geoghagen,N.S., Weidman,J.F., Fuhrmann,J.L., Nguyen,D.T., Utterback,T., Kelley,J.M., Peterson,J.D., Sadow,P.W., Hanna,M.C., Cotton,M.D., Hurst,M.A., Roberts,K.M., Kaine,B., Borodovsky,M., Klenk,H.P., Fraser,C.M., Smith,H.O., Woese,C.R. and Venter,J.C.				
TITLE	Direct Submission				
JOURNAL	Submitted (27-AUG-1996) The Institute for Genomic Research, 9712				
COMMENT	Medical Center Dr, Rockville, MD 20850, USA				
FEATURES	On Jan 30, 1998 this sequence version replaced gi:1591709.				
source	Location/Qualifiers				
	1..14561				
	/organism="Methanococcus jannaschii"				
	/db_xref="taxon:2190"				
gene	184..1056				
	/gene="MJ1057"				
CDS	184..1056				
	/gene="MJ1057"				
	/note="similar to percent identity: 31.93; identified by sequence similarity; putative"				
	/codon_start=1				
	/transl_table=11				
	/product="glycosyl transferase, putative"				
	/protein_id="AAB99061.1"				
	/db_xref="GI:2826364"				
	/translation="MVGEMDKPLVSVVMTYNEPEKYLKESIESIXNQTKDFYIIVLDPNPKAEEIKEYOOKRKRIIFIKERNLGRGASRNKAVNLARKYITAILDADDIALPKLEKQFKYMNNRDIIDLLFSWYFIDENGNIILKEPKFKYKFKYKFKHLTVHPMMVKSKILKLYKDEKILRSQDYDFWIRCIANDYKFDIIEEFLKRYIPNRNLYLSRIKKOKLYSYVTLTATHWNKKHFCNNVYFWKVFVSLVVLIVLPTTFILKILI DIKKDKTEISTKGH"				
	1067..2896				
gene	/gene="MJ1058"				
CDS	1067..2896				


```
gene
/product="Gmd"
/protein_id="AAL61891.1"
/db_xref="GI:18152900"
/translaton="MKKALISGCTGQDSGFLAEFLQKGYEVHGLIRRSSEFNTRGRIE
HLVDEWVRDMKQKTVNLHVGDMDSSSLIRIQQVOPDEIYNLAOSHVKVSEFDP
EYTAADALGTLRLMEAVRIILGLEKQTRIYQASSELGKQVEPQSETPFPYRSPY
GVAQGYWITKNYRESYSGMAVNGILFNHESERGETFVTRKHTLAAARIAQGEQDK
LYGLNDLAKDMGAKDZVECMFLILOHDPEDFVIATGEMHTVREFCTLAFAEIGIN
LRWEGEVNKGIDTATGKVLVEVDPKYFRPAEVBQLLNGNPTKARTVLGNPCKTFFP
ELVKLMVSHDMAKVRMTATKHD"
5702..6973
/feature="Orf3"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAL61892.1"
/db_xref="GI:18152901"
/translaton="MNDIFDELVSNSWNGKLPKVEFARKSYTNRFYVGNRLVKVL
TGQGVGKGLQLIQLIADGVSSRNILYSKFTDSGLSGTRDIDDDLLFIYREK
IHPGVKVIYFEEIQKVEGWEHFVSHSQDFVDTCELFISGSDMLSGEAEKLLARH
VYSFIFPFSFSEHRIERLEVEATGKNYAEYMGKALPILFNLPAEAGKWAISALRDT
ILFRDIIRYRIKARLLEDVFLYCTHIAELITIGQLVSHAQNRKTSYDTVANYI
CYLEPISLHVERVOIRSKETLLGSKYIINDWSFMRHLYPFTAGSRKSCFNOVL
ELRRAGYTVFVGLRGKLVDFVGRKKDRIVYLOCAPLLNDDFQVEQWYNLEMLIQDN
YEKVVVSMDDTTLPSKEGIRHIQWOLPEIL"
complement(7267..8070)
/feature="Orf4"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAL61893.1"
/db_xref="GI:18152902"
/translaton="MATTYDGINPPIAVNFEWNAMEVIEAKYMGKGLAIVLKLLCK
IVKEYYIYPWCEOCLIFANKTGKEADEEVOGIIEMIEKGMIDRGSYAKHVLVTS
AIQKVMEATRRKRNWTFAMPYLILIKPKETCNEKTVCTQNVQDAGSDAKNACNTEQ
SKYKOSREFPSAPRKEEVNATPVSMGPFATFNMTNIVGLMGNLERFGITDEKE
IEAILRLSDYGRKGPVWKLICSTNWSNIGAKGYMIAALNRAKRKRSCT"
complement(8120..8467)
/feature="Orf5"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAL61894.1"
/db_xref="GI:18152903"
/translaton="MKKRNKAVSKNEMMKWKRMQIMLEKDYTEQCSEWMAERLEALI
EYMQYGHAAVAYMKODGTFKLVKGTLVGYEKDFGQYDPMPEIKNTVYVRDVEQQRWMT
FKIENFEWRAIV"
9468..9992
/feature="upb1"
/feature="upb2"
/feature="upb3"
/codon_start=1
/transl_table=11
/product="putative transcriptional regulatory protein"
/protein_id="AAG26461.1"
/db_xref="GI:11023504"
/translaton="MSEQKYWFARTRDKQFEAIRDSLEKLTDLNLYLPTQFVI
ROLKYRRKRVPEVPIKNIIFQATKQDADCSNKYNIQLFYMKDLITRAMLIVDPKQI
QDFIVMDLPNGVSFDDNHLVSGSVQVVKGDFGCGVSEGLASEANKYVVRIRAGVL
SASVKPKSYLRVI"
9996..10481
/feature="upb2"
/feature="upb3"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAG26462.1"
/db_xref="GI:11023505"
/translaton="MNTLTSTQEQSLAHELILYGVGAPLYTDFHFRQLNKEVLEQS
DALXPQRGATPEEANICLALMGYNATYINOGDKEKKQVVLNRCWDLVDQLPATLL
KCOLLYTCYGEVFEELAKEAHTIIESNSNRELLKAEKIAESLNLEANPYPSLEH
E"
```

```
gene
10478..11761
/feature="wcf1"
10478..11761
/feature="wcf1"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAG26463.1"
/db_xref="GI:11023506"
/translaton="MIRFSLRSPLYLALVSLIFLVYTCAYARKPFTAAIYEGE
ILMGPDVKILYSELIGYALSKFIVGILPSMKAGHRIYVIGLITSEVALLGFLAL
LPVPLKVCISIFLGPLGIMWGVIFSYIEGRISSEILNGLVSLVSSGLVKLTGQF
VMDNLHVSEYWMPTTGLLVYPMMLCSWLLNIPAPNEQDIIQRTKRAPMDKDRRK
FLRFFWGCIMLVVYFALTFRDLRSDFAADINWELHEGAMIFQTEPVTAAAFVLV
LMFLIVFVRNRLNLIYCIATVGTGSLMVFSTLLYVYGFGLSPIWMMILSGGLYMGY
IPFTYLIERLIASLRVSTAVFIYIYADSFVGLTGTGVFMVKNFVSADISWTVMLMRT
ALFSGFVSVLSIFAIYCYFKQQLNSLILIPSEIHD"
11754..12473
/feature="wcf0"
11754..12473
/feature="wcf0"
/codon_start=1
/transl_table=11
/product="putative glucose-1-P-cytidylyltransferase"
/protein_id="AAG26464.1"
/db_xref="GI:11023507"
Query Match 12.9%; Score 120.8; DB 1; Length 34508;
Best Local Similarity 60.2%; Pred. No. 5.8e-10;
Matches 200; Conservative 0; Mismatches 132; Indels 0; Gaps 0;
QY 4 aattatgatatcattatcggttatataatgagcccttaataattatgagagattcagta 63
DB 22262 GAAGATTAGCGTATGATGCTATATATAGAGAAATCTTTATGTAAGACAAGCTATT 22321
QY 64 gaattatattaaatacaacgcttactgattttgagttgcataattgttcattgataatcca 123
DB 22322 GATAGTATTTAAATCAAAAGTTTGGAGATTTTGAATTTTAAATATATAATGATAATCCA 22381
QY 124 agtagaggtgatttaagaacattcttaacagaataattcagctgtagataataagaataaa 183
DB 22382 GAAGATTTTATTTAGATAAGTTAGTTTGAATAATAGCAAGAAAGATAGTCGTATACAA 22441
QY 184 atctgtcctaagaagaaataattggttttagcatcaagtttgaacaaacggtgaaaaatt 243
DB 22442 TATCAACGAATAAGAAAAATATAGTGTAGCAGCCACTTTGAAATCGTCAGATTTTGTGA 22501
QY 244 tctaagggagaataatatttttagaataagtgatgctgtagatatttcataatccaagtagattt 303
DB 22502 GCTAAGGTGAGTATTTAGTCTGATGATGATGATGATATTTTCAATACCGGATAGATT 22561
QY 304 gataagcaaatctcgtttttatggaggagaaattc 335
DB 22562 ATTATGACGTTGAATCTAGATAGATAATAC 22593
RESULT 5
AE007717/c
LOCUS AE007717
DEFINITION Clostridium acetobutylicum ATCC824 section 205 of 356 of the
complete genome.
ACCESSION AE007717
VERSION AE007717.1
KEYWORDS GI:15025159
SOURCE Clostridium acetobutylicum.
ORGANISM Clostridium acetobutylicum
Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
Clostridium.
REFERENCE 1 (bases 1 to 14157)
AUTHORS Nolling,J., Breton,G., Omeichenko,M.V., Markarova,K.S., Zeng,Q.,
Gibson,R., Lee,H.M., Dubois,J., Qiu,D., Hitti,J., Wolf,Y.I.,
Tatusov,R.L., Sabathe,F., Doucette-Stamm,L., Soucalle,P.,
Daly,M.J., Bennett,G.N., Koonin,E.V. and Smith,D.R.
```

TITLE Genome Sequence and Comparative Analysis of the Solvent-Producing
Bacterium Clostridium acetobutylicum
J. Bacteriol. 183 (16), 4823-4838 (2001)
MEDLINE 21359325
PUBMED 11466286
REFERENCE 2 (bases 1 to 14157)
AUTHORS Childress,D., Zeng,Q. and Smith,D.R.
TITLE Direct Submission
JOURNAL Submitted (24-JUL-2001) CTC Sequencing Center Production,
Finishing, and Bioinformatics teams, Genome Therapeutics Corp., 100
Beaver Street, Waltham, MA 02453-8443, USA
FEATURES
source location/Qualifiers
1. .14157
/organism="Clostridium acetobutylicum"
/strain="ATCC 824"
/db_xref="ATCC:824"
/db_xref="taxon:1488"
complement(138. .1136)
/gene="CAC2166"
complement(138. .1136)
/gene="CAC2166"
/codon_start=1
/transl_table=11
/product="Nucleoside-diphosphate-sugar epimerase"
/protein_id="AAK80124.1"
/db_xref="GI:15025160"
/translations="MNLVTGGAGFIGRWVVKLLDDGNTVTALDNLNGSIEINIEF
KGNFPIEGDIKNKDDLDRIEDKYDIYVHLAASIVVQESIDDPKPTFNDVTGTFN
ILERAKQMGQHGKNGSINWIDENDEYVPCKVPMSTWYVDISGETGIDQHSVK
PVSPYGSKTAAMNVLVSFYNAKLPVIRPNTYGPFOKSGEGVSIFLKNFID
NKPINYSKQKTRDLFLVKDCADFYVNAVGYKSEVDGEIINAGTGNDITINELANLIA
GKVSYNHVHGHIPQSEIMKLCNYSKAKNLGMGWEHYTLEEGIAETKSWLASTIK"
complement(1268. .2119)
/gene="CAC2167"
complement(1268. .2119)
/gene="CAC2167"
/codon_start=1
/transl_table=11
/product="Flagellin family protein"
/protein_id="AAK80125.1"
/db_xref="GI:15025161"
/translations="MRINHNMAALSIFYEQKSVGGQATSLERISSGLKINSKADGPA
ALBKSQMRNQIRGMMAARNVQDGISMFQTAEGGMSNVSDMLQMRDLVLESGDSK
TSDKKNIOLEIKNGINDITNTNFNGKPLLASNEKSEITIGNNVGETVOFSFYD
LSSESIGOKDANGVTSIRNVNLOADSLNANLATIDSALSTVLDARAQYGAENRPE
DTYNLNSPESGVQQAESDIRDSVADEMITYTKTGVLMNAGDAMLAAQNLPKDILN
ILSIRK"
complement(2227. .3963)
/gene="CAC2168"
complement(2227. .3963)
/gene="CAC2168"
/codon_start=1
/transl_table=11
/product="Uncharacterized conserved protein"
/protein_id="AAK80126.1"
/db_xref="GI:15025162"
/translations="MNIITFNTVMGCRDISEINAVIEESLDNKNILKDNKYGNREN
VQEDINSFKKLNIENTESDFVIFGLGCGEHLKLLNMNLKMKRLILVEPREDVISRF
LKLNYANKIIDDRIVLYLYNEEDLNCILSLFDENISNTKCGVFANYADTYEELS
NLKSYDIQNSLLVNMGTHIVHSRHFSSYMMNLKFIQKSTIINHFNKIKGMPAVI
VSAGPSLNKLDLDYQSKFIVITGGRTKLKLLDKGIRPDFVCVIDPDEPAFEIMKD
ALDCTVP IYSEFTNRYKIKYKDKIFFTDTGVNDASDKDFEKKIDNLYEGGSVAHV
CASLAEY IGCNRIIFVQDFAYTNDVTCDLNVGNSNIKNFYVDDVYGGKVKTDRTL
YFYKMAEEFIOVKKIVDFINCTEGGANIEGAPLIKPKDALEKGEISIKNTEAVSNII
KSNVLMDDNMRIKLDGIRKQNIKVISSMCTSAVEQYNSYLAFGKSESOKFLAHNF
NLNNECIFSNASSSFKFINVLSPAILNIGLNPREFKMGKMBEKLKIKPERNLILYN
NIIEAAKEFNKY"
complement(4024. .5106)
/gene="CAC2169"
complement(4024. .5106)
/gene="CAC2169"
/codon_start=1
/transl_table=11

/product="UDP-galactopyranose mutase"
/protein_id="AAK80127.1"
/db_xref="GI:15025163"
/translations="MEDYLVGCGFCGAVSARLLAEKGGKILLIVDKRHHGGNAYDIY
DNGILVHKYGPPIHFTNRSRVRKDEIFLRFRTKWIYQHKVLSYDGRVLVPMPIINDVYN
OLYGTNYNDTIIEFFDKVREKDEIINNARDMVYTKGSLYKMFYTKKONGIMP
EELEKEVTARIPITNRDRDYFTNYOGLPLDGYTKMFENMLNHNHILMLTYKFI
KSEIQYKSLIYTGCADEFDYKFGRLPYRSINFETLNODYVOKGVTVNYPNDYDFT
RITFEKHLTGOKCYNTTIVREYSCSDGEYYPVPQSKNQKLYLEYKKQVENLENAYFL
GRUGOYRYMNDVVVDEAIKLVOKLK"
complement(5153. .5638)
/gene="CAC2170"
complement(5153. .5638)
/gene="CAC2170"
/codon_start=1
/transl_table=11
/product="Predicted membrane protein"
/protein_id="AAK80128.1"
/db_xref="GI:15025164"
/translations="MKRLIVRSASFQOOLDNLKAIKKYKYDEIDILTHEHGKVLAE
KYNNIHVIYVPYKSEFKGKVKKEIKNFVDVIVPVTNISGAGFNVFKYALSIINS
DKIIMCNVISEFKDRDRLSIYIMELSSFVKILSGFTCLFGIVSVFLIFLKLRAAK
K"
complement(5661. .6560)
/gene="CAC2171"
complement(5661. .6560)
/gene="CAC2171"
/codon_start=1
/transl_table=11
/product="Predicted glycosyltransferase"
/protein_id="AAK80129.1"
/db_xref="GI:15025165"
/translations="METPLYSIVSICWNKRKEICESLNNIRKIDYESLEIIVDNGST
DCTVEKIEEDFKVKLLIKMPKNAGIEAYNVGFKSAAGEYIVILDDDFPEPKNSIKRW
KKFQKDDKLGWAFDVRNRYNIDEVKIDITEETNSDAAESETIMAFNGAGAVRTE
VLKAGFYPEEFYLYWNEQDTAFRLDMGYKIFKFDVIAIYHKYSPKNRASWRAPFY
TRNAFFLVKNYSTKAALKKTIELIYDCFYYSMEQKNTIYLKAMWNAFMNFNKIKGR
KVVDKYIEENLRIPLVNSFTFYK"
complement(6689. .8038)
/gene="CAC2172"
complement(6689. .8038)
/gene="CAC2172"
/codon_start=1
/transl_table=11
/product="Predicted glycosyltransferase"
/protein_id="AAK80130.1"
/db_xref="GI:15025166"
/translations="MKELSIVCNKNKEYVNLICDSVLNSSYRDLDIYVVVDNASEDG
SVEAIKDKFKDVKMLISNTENIGSGGFNTGIREVLKDKYKYMLLDNDVILDKKAIN
NSYRYLEKNDDTAVVGSKIYSMDNRDKIQEMGAKIDFENFYINPFYGYIDFIELPEV
VEGDYVPACSMVRIEALRRVIMDEDNFIWDDIENGWYRFKAGYKVTYGGKSVWH
KMGVAOKTNTFGTYFWRNRIREFAKYIEENRKIDFCILRFDEIFOALYSCNYIGKYS
SARTIIAAVEDALNDVRGKASQDKILIEIKVKDRFGLIENKKNKIALTGDSNIEVLN
IVNKAITNSETELNLSGNVKNIKQQFEYKVLKSESLKTYETVIKICNHFDMR
NELSNLIYVDKFPNVIVSGDDRKYIENYDNTYSLMKNIWPNLLSKMRVNLRLNKK"
complement(8063. .9064)
/gene="CAC2173"
complement(8063. .9064)
/gene="CAC2173"
/codon_start=1
/transl_table=11
/product="Glycosyltransferase"
/protein_id="AAK80131.1"
/db_xref="GI:15025167"
/translations="MAKVSVMVPVYNSEKYLKESIESILNQSYSDLEFIINDGSTDN
SFKTIKLYAKDKRINVISRENKGIYSLNAEIRLAKGEYIARDADDISAPKRIEQ
ISFLKSHRDILGTQVYVNGSINIKENKNLNIEFDIYDNRREKILNMYWYCLAH
PSVMFRKIDILRELKGYNDFKSEDLDLWLRATESGFKYIKLEEILYFMHSESRTVD
NONYEGLDGKIKLIDVFKREFKDFKIYVWNGSKGKITKVELDFFEFKSOCIAFY
DKFTGFEKIRILHPRKIDNIFDYVFIATEPKKEAMSSLSKGLKCKDKFLSTV"
complement(9115. .10125)
/gene="CAC2174"
complement(9115. .10125)


```
/gene="CAC2174"
/codon_start=1
/transl_table=11
/product="Glycosyltransferase"
/protein_id="AAK80132.1"
/db_xref="GI:15025168"
/translation="MKVYPKVSVMPVNSERYLAETIESILDQTYNDFEIIIVDDGS
TDESNISSVANKDNRIIVISREHGLVDSLNEGINSRARGKIARMDDADISINRI
EKQFEFLKNDVDILGTRIEAFGDIDEKQTIYNSAFSIFKSDNISQVFLTSKAIP
HPSVMFKLSDSVKLGRYKEDYADYDLMLRAIRNGKIVRMDDECLKYRVHNSKT
AVENPKWVEYTKAKIDYINDTNKDKVDYLWGASTGKLVKKVVESTYDKFNLI
GFIDYKEGEVDECIYKPEDMLKLSHDYVFATLPGKEADEFLKNCGLKFKFEYIS
Lv"
gene
complement(10149..11150)
/gene="CAC2175"
CDS
complement(10149..11150)
/gene="CAC2175"
/codon_start=1
/transl_table=11
/product="Glycosyltransferase"
/protein_id="AAK80133.1"
/db_xref="GI:15025169"
/translation="MISVIMPVNCEKYLESSIESILKQTYRDEFEIIIVDGSNDKSI
DLINKYANDNRIVVSRDNNMGWYISNEGIDRAKGSIVARMDDADIALPERFERQI
EYLNKNDVDILACKVEAFGDVSRQKLEREHWNVDLSSESIESLENCYTAHPS
VMKMSVLKALGGYNLKRTEEDYNLRAIAKGYIAMLEEKLMKIRLHNSDKTHRD
AEGSSIRDIQSRLEYKELKLDKLVYIWGASNGGKIAYEKIKEYFNPAKNGYI
DKFTGKPEGVDFIYFEDLSKMKVDYVFATLPGCKEAKGILTKNFPVINKDLVLC"
gene
complement(11189..14002)
/gene="CAC2176"
CDS
complement(11189..14002)
/gene="CAC2176"
/codon_start=1
/transl_table=11
/product="Glycosyltransferase domain containing protein"
/protein_id="AAK80134.1"
/db_xref="GI:15025170"
/translation="MIKVKDFVSFIERSRNLIENPNSVILCLINIKDYNKLDKSI
SLILNQYRNFELIVDDQDBELLYTKIDYINKDNRIYVYVHNNRSLAALRINQA
Query Match
Best Local Similarity 50.1%; Pred. No. 1.9e-09;
Matches 467; Conservative 0; Mismatches 440; Indels 25; Gaps 6;
QY 13 atcattatgcgtatataatgagcctttaaatattgtgagagattcagtagaatactata 72
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 9052 AGCGTGTGTAATGCCAGTTTATAATAGTGAATAATCTCGAAGAGTCCATAGAAAGTATT 8993
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 73 ttaatacaaacgcttactgatttgcgtatcattgtagatcattgtagatcattgtagatgaggt 132
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8992 TTAATCAATCATACAGTGTGTAATTTATAATAATAAATGATGGCTC---TACAGAT 8936
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 133 gatttaaacgaattcttaacagaattcagttgtgtagataataagaataaaattcttgcct 192
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8935 AACTCTTTTAAATAATAAAGGAATATGCCAATTAGATAAAGAAATAAA---TGTGATT 8879
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 193 aatgaagaaattgtgttagcatcaagtttgaacaaagcgggtgaaattttctaaagga 252
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8878 TCAAGAGAAATAAAGGAATTCGTACTCGTTAAATGAAGCTATACGATTGGCAAGAAGA 8819
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 253 gaataatttttagaatgctgtagatatttccatcccaagtagattgtagaagcaa 312
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8818 GAATATATAGCAAGGATGCTGATGATTTTCGCCACCCCAAAAGAAATAGAAAAACAA 8759
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 313 attcgtttttatgga---ggaattcattggtttctcagcaactcctaataattgata 369
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8758 ATTCTTTTCTTAAAAAGCCACAGAGATATAGATATACTGGGACACAGCGTAAAGTGGTG 8699
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 370 gacccaaaaagaaatttagtatataaacacagagaaagtaataaaaa-----ta 417
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8698 GCGAATATTCTTAATGATATAAAGAAAAAATAAGCTAAATATAGAAATTGAT 8639
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 418 tacttaactaargatatacaggaagattgttattgaatagatctatacttgcaccccaacg 477
```

```
Db 8638 ATTTATGATGATAAATAGAGAAAAAATATTAAATTTATGTTGTTAGCACATCCTTCA 8579
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 478 tgggtcgtaaaaaaagaagtcttcgataagatttaaggatagagatttagtacgtgt 537
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8578 GTAATGTTTGGAGAAAGATATATTGAGAGAGCTATAAGAAAGCGGATTTAAATATATATAAGCTAAAG 8522
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 538 gaagattatgattttcgaataagagagagctctggctgatttcaaaaatcggcttactcaat 597
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8521 GAGGATTTAGATTATATGCTTAAGAGCTATAGAAAGCGGATTTAAATATATATAAGCTAAAG 8462
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 598 aaagtacttttaccagtatagattaaacgagaatgggaatatcacaaacccaataagtttaag 657
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8461 GAAGAACCTTATATATTTAGGATGCGAGGAGCTCAAGAGCTAGAGTAGATAATAATAAAT 8402
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 658 caatatatttactcagctattttacaagatttttataaagaaaaaatcttatattgataac 717
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8401 TTGAAGGTTTAAAGATGAATAAAAAATAAATAATTAATAGATGTTTTTAAGAGAGAGTTC 8342
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 718 acaaaaattactaattacttctcaagagatgtgataaagaacgctatactcagcagaag 777
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8341 AAGAAAGACTTTAAATACATAGTATGCGGTGCAAGTAATGGAGGCAAAATTCACAAAAGAG 8282
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 778 ctctctaaatattttgagtgataaatactacccctagtagtatttactattagaaacatatatt 837
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8281 GTTTTAGAGCGAATTTTTTTGAGAAATCTCAATGCTATTTGTCATTTCTAGATAAAATTTAAACA 8222
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 838 tgtttattttacttacttgaagctctcc-cttggttagagggtatttaataatgatattaa 896
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8221 GGTGAATTTGMAAAATTAAGATTTTCATCCAAAAGATATAAATCATATATAAATTTGAT 8162
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 897 tattttagtactgaaattgtttggaggagaga 928
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8161 TATGTTTTTATTGCAACTGAACCGGAAAAAGA 8130
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
RESULT 6
AF279134 3584 bp DNA linear INV 05-SEP-2001
LOCUS Dictyostellium discoideum alpha1,2-fucosyltransferase (fuch) gene,
DEFINITION complete cds; and unknown genes.
ACCESSION AF279134
VERSION AF279134.1 GI:9022424
KEYWORDS Dictyostellium discoideum.
SOURCE Dictyostellium discoideum.
ORGANISM Dictyostellium discoideum.
REFERENCE 1 (bases 1 to 3584)
AUTHORS van der Wel,H., Morris,H.R., Panico,M., Paxton,T., North,S.J.,
Dell,A., Thomson,J.M. and West,C.M.
TITLE A Non-Golgi alpha 1,2-Fucosyltransferase That Modifies Skp1 in the
Cytoplasm of Dictyostellium
J. Biol. Chem. 276 (36), 33952-33963 (2001)
JOURNAL 21423935
MEDLINE PUBMED 11423539
REFERENCE 2 (bases 1 to 3584)
AUTHORS van der Wel,H. and West,C.M.
TITLE Direct Submission
JOURNAL Submitted (15-JUN-2000) Anatomy & Cell Biology, University of
Florida College of Medicine, 1600 SW Archer Road, Gainesville, FL
32610-0235, USA
FEATURES
source
Location/Qualifiers
1..3584
/organism="Dictyostellium discoideum"
/db_xref="taxon:44689"
<1..>807
/product="unknown"
<1..807
/codon_start=1
/product="unknown"
/protein_id="AAF82377.1"
/db_xref="GI:9022425"
/translation="DLKQFDLSLNQYCDENILSTYILGNACPRWTPPHQGHNRNNSQ
```


QQLYPSANPQOQOPLIPPTILSQPOKYTPPTNHHHHHLLNCLSPRISYNTLSPTQOP
INTSVNNNNIIQORNSISSSNCINIPSSRRSFSDNIINGNGNNGVNGVNGSSSLS
NTTSAFSPNVNSYSGNSGSCNISSSMSMMMSGGNQOPLIGSPSEIMHKLKE
IERLNKEATQKQOQELNQVKIAEABERSLCLDKYKEYRNELIERELVK"
<1067..>3373
/gene="fucB"
/product="alphal, 2-fucosyltransferase"
<1067..>3373
/gene="fucB"
1067..3373
/gene="fucB"
/note="glycosyltransferase"
/codon_start=1
/product="alphal, 2-fucosyltransferase"
/protein_id="AAF82378.1"
/db_xref="GI:9022426"
/translation="MNDSPISVLPFLIKNDKSLNYQGINLLIISIDSIIIEOTFK
ENELIVDDGSNEIEQLSKRYSDNRKIFLTKENKGIYKSLNDALINHCSPTSK
YIARMSDDISHPTROSQLYQLOSHETIDILGCPKMFNNKLLIEILNNNNNNIN
NAVKELDIIINNESEFKIQHPDKDILMSMFNCCIVHPSVIFKRSIFTIEHYEEN
NQPFTELYLFWLKLISLINSQSSPTLLYLKRNHNSISFKNIEKODSTANAS
CYVNLTPKREINDSEIIIONSLSMKEIIQFOLSPSSLSKINNIIEIPEFAFKYLE
LLEKSTCKQOQVNSIKDAANEKKGELYSLCLSNYPNNQKSSLLKEKLSRNPSTQL
LSLLSNLVSSTTIINNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNLNFIGIN
SNKINTPPKNNNKKFENGIRIICFSKDRAPQLKEYLRTFFKYLKNDNDNDKFEIIVD
VLFYTSNERFKNSQVIESFPQVNFIKENFTDQINLVQKTNKLEYVMFSYDDILY
YNEFNLEYCLSLNSPLALGFYMKLNKNTYCHTQNOQITPLNSNTISRTENNFKY
LKNRNDKCDKWNYPWDLCTIYRCNDIDSTINGIVKYGIRNGINPNRFEENG
RPIQKQIYQNKPYCLCLSDHSPNSVSTINRVQDYVDNPIDYDTISLDDLDLQLYSN
KSLNDKRYENSLNFKSVHIGELFIS"
complement(<3471..>3584)
/product="unknown"
complement(3471..>3584)
/codon_start=1
/product="unknown"
/protein_id="AAF82379.1"
/db_xref="GI:9022427"
/translation="DPTKENNTVGHVYKOGYRUHLDRVPMVGVNKRKPQ"

BASE COUNT 1487 a 442 c 365 g 1290 t
ORIGIN

Query Match 10.9%; Score 102.2; DB 3; Length 3584;
Best Local Similarity 50.3%; Pred. No. 6.7e-07;
Matches 285; Conservative 0; Mismatches 273; Indels 9; Gaps 1;

Qy 1 atgaattatagtcattatgcggtataataagagccttaataattatgagagattca 60
Db 1109 ATAAGGACAAATGACGATAATATTAAATACCAAGCAATAATAATTTAATAATATCA 1168

Qy 61 gtagaactctattataacaaagcctcactgatttgagtcataaattgtcattgataat 120
Db 1169 ATAGATAGCATTTATGACCAAACTTTTAAAGAAATGGGAATTAATTTAGTTGATGATGGA 1228

Qy 121 ccaagtagaggtgatttaaagcaattcttaacagaataattcagttgttagataatagaata 180
Db 1229 TCAAAATAATGAAATTTGGGCAATTAATCTTCAAAAAGATATAGTACAGATAATAGAATT 1288

Qy 181 aaaaactctgtaagagaaaattatggttttagcatcaatgattgaacaaagcggtgaaa 240
Db 1289 AAATTCATAATAAATAAGAGATAAAGGTAATGTTAAAAAGTTTAAATGATGCAATTTTA 1348

Qy 241 att-----tctaagggagataatatttttagaagtgatgctgattgatttcattat 291
Db 1349 AATCATTTGTTCAACCAACTTCAAAATATATCTCTCGTATGGATTCAGATGATATTTCTCAT 1408

Qy 292 ccaagtagatttgaagcaaatcgttttatgaggagaaaattcattggtatttcagca 351
Db 1409 CCAACAAGATTACAATCTCAACTTAAATATCTTCAATCAATGAACAAATTGATATATTA 1468

Qy 352 actctaataagattgatagaccaaaaggaattatagatatataaacaacagagaagaataat 411
Db 1469 GGTGTGCCAATTAATAATGTTTAAATAATAATAATTAATTGAATTTTAAATAATAATAAT 1528

Qy 412 aaaaataactaactaatgatatacggaaagatgttattgaatagatctatacttgccac 471
Db 1529 AATAATAATAATAATAATAATAATGTCGAAGAGTTAATAATAATAATAATAAGAGAA 1588

Qy 472 ccaacgtggtgctgcaaaaaaagaagtttgcgataagtaataaggatagagattagta 531
Db 1589 TCITTTAAATTTTATTCACATCCCTGATAAAGATATTTTAAATGTGTCATTTTTCAT 1648

Qy 532 ccgttgaaagattatgattttgcaata 558
Db 1649 TGTGTATTTGTTACCCCTTCGTGTAATA 1675

RESULT 7
AP001519/c
LOCUS 303650 bp DNA linear BCT 10-JAN-2001
DEFINITION Bacillus halodurans genomic DNA, section 13/14.
ACCESSION AP001519 BA000004
VERSION AP001519.1 GI:10176109
KEYWORDS
SOURCE Bacillus halodurans DNA.
ORGANISM Bacillus halodurans
Bacteria; Firmicutes; Bacillus/Clostridium group;
Bacillus/Staphylococcus group; Bacillus.
REFERENCE 1 (sites)
AUTHORS Takami,H. and Horikoshi,K.
TITLE Reidentification of facultatively alkaliphilic Bacillus sp. C-125
to Bacillus halodurans
JOURNAL Biosci. Biotechnol. Biochem. 63, 943-945 (1999)
REFERENCE 2 (sites)
AUTHORS Takami,H., Nakasone,K., Sakiyama,T., Maeno,G.,
Sasaki,R., Hirama,C., Fuji,F. and Masui,N.
TITLE Genetic analysis of the chromosome of alkaliphilic Bacillus
halodurans C-125
JOURNAL Extremophiles 3 (3), 227-233 (1999)
REFERENCE 3 (sites)
AUTHORS Nakamura,Y. and Inoue,A.
TITLE An improved physical and genetic map of the genome of alkaliphilic
Bacillus sp. C-125
JOURNAL Extremophiles 3 (1), 21-28 (1999)
REFERENCE 4 (sites)
AUTHORS Takami,H., Masui,N., Nakasone,K. and Horikoshi,K.
TITLE Replication origin region of the chromosome of alkaliphilic
Bacillus halodurans C-125
JOURNAL Biosci. Biotechnol. Biochem. 63 (6), 1134-1137 (1999)
REFERENCE 5 (sites)
AUTHORS Takami,H., Takaki,Y., Nakasone,K., Hirama,C., Inoue,A. and
Horikoshi,K.
TITLE Sequence analysis of a 32-kb region including the major ribosomal
protein gene clusters from alkaliphilic Bacillus sp. strain C-125
JOURNAL Biosci. Biotechnol. Biochem. 63 (2), 452-455 (1999)
REFERENCE 6 (sites)
AUTHORS Takami,H.
TITLE Genome analysis of facultatively alkaliphilic Bacillus halodurans
C-125
JOURNAL (in) Extremophiles in deep-sea environments (Ed.);
HORIKOSHI, K. TSUJII;
: 249-284; Springer-Verlag (1999)
REFERENCE 7 (sites)
AUTHORS Takami,H., Nakasone,K., Ogasawara,N., Hirama,C., Nakamura,Y.,
Masui,N., Fuji,F., Takaki,Y., Inoue,A. and Horikoshi,K.
TITLE Sequencing of three lambda clones from the genome of alkaliphilic
Bacillus sp. strain C-125
JOURNAL Extremophiles 3 (1), 29-34 (1999)
REFERENCE 8 (sites)
AUTHORS Takami,H. and Horikoshi,K.

TITLE Analysis of the genome of an alkaliphilic Bacillus strain from an industrial point of view

JOURNAL MEDLINE Extremophiles 4 (2), 99-108 (2000)

REFERENCE 20263314

AUTHORS 9 (sites)

Nakasone,K., Masui,N., Takaki,Y., Sasaki,R., Maeno,G., Sakiyama,T., Hirama,C., Fuji,F. and Takami,H.

TITLE Characterization and comparative study of the rrn operons of alkaliphilic Bacillus halodurans C-125

JOURNAL MEDLINE Extremophiles 4 (4), 209-214 (2000)

REFERENCE 20426005

AUTHORS 10 (sites)

Takami,H., Nakasone,K., Takaki,Y., Maeno,G., Sasaki,R., Masui,N., Fuji,F., Hirama,C., Nakamura,Y., Ogasawara,N., Kuhara,S. and Horikoshi,K.

TITLE Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and genomic sequence comparison with Bacillus subtilis

JOURNAL MEDLINE Nucleic Acids Res. 28 (21), 4317-4331 (2000)

REFERENCE 20512582

AUTHORS 11 (bases 1 to 303650)

Takami,H. and Takaki,Y.

TITLE Direct Submission

JOURNAL Submitted (22-MAR-2000) Hideto Takami, Japan Marine Science and Technology Center, Deep-sea Microorganisms Research Group: 2-15 Natsushima, Yokosuka, Kanagawa 237-0061, Japan

(E-mail:takami@jamstec.go.jp,

URL:http://www.jamstec.go.jp/jamstec-e/bio/DEEPSTAR/FResearch.html,

Tel:81-468-67-3895, Fax:81-468-66-6364)

FEATURES
source

1..303650

/organism="Bacillus halodurans"

/db_xref="taxon:86665"

/note="alkaliphile"

complement(103..1887)

/gene="BH3486"

complement(103..1887)

/gene="BH3486"

/codon_start=1

/transl_table=11

/product="butyryl-CoA dehydrogenase"

/protein_id="BAB07205.1"

/db_xref="GI:10176110"

translation="MGNTAQTGTVKSGSLDEIDADKVFTPEDFTEHEMIGKTTEEFVYNEVPVLDIEHNFQDISRRLLTQAGELGLGADVPEYGGGLDKRISSTIIEKF SRAGFSLSYGAHVIGSLPIVFFGNEQDKKYLPLDLAGARIAAYALTEPSSGSDAL GAKTTAVLINEAGTHVLNGEKWITNSAFADVFIYAKIDGHEFSAFIVEKDFEGVST GPEKKMGKSGSTFTLLEDALVPKENLGEVCKGHVIAFNILNVGKYLGVGCIIG SKRGIELAAKNAERKQKQPIAKFTLQEKLANMAVITYAAESSIYKGTGLFEDRLG GLSEEQKDGREVAKAIAEYALECSLNKVGSEALDFVADEAVQIHGGYGFMAETEVE RMYRDSRINRIFESTNEINRLVPGTILRKAMGELPELEKATALQEEMLMFPQEVG DEPLEQERYLLSMARKVFLMIAGTGAQTYGKQLQEQELLANVADIVSDIFSMESVIL RTEKAIRRNGLEKAEQKLTQVFCQEA FNIEAHAKESLSVAMQSGDLTRMTSTLRK LTRHTPINVIKKREIAGRIEAERYV"

complement(1909..3090)

/gene="BH3487"

complement(1909..3090)

/gene="BH3487"

/codon_start=1

/transl_table=11

/product="acetyl-CoA C-acyltransferase"

/protein_id="BAB07206.1"

/db_xref="GI:10176111"

translation="MREAVIVAGARTPVGKAKNGTLANVRPDDLGAITVKETLRAGD FPARIDDIIVGICAMPEAESMSLI PMGGHVIAFNPTLVENQPEYVMGMGHTAEVASRFG IRIMLQGEALVAGAESMSLI PMGGHVIAFNPTLVENQPEYVMGMGHTAEVASRFG VSREQDAPAVESHKRAAKAIEEGKFADEIYVPDVTLRHVDNHLKVEKHVTFSDVEG VRPEINMETLSTRPFAHAKGTVTAGNASQMSDGAVALMDREVAEREGLTPVVKFLR SFVAVGAPVEGVGPEAIPKAVMAGLDLSDTGLFELNEAFASQATQVIGRLNLDH DKVNVNGAIALGHPGLGTGTKLTLSLHHEMKRREGQFGVVVTMCIGCGMGAAGVFELI G"

complement(3111..5492)

/gene="BH3488"

complement(3111..5492)

/gene="BH3488"

/codon_start=1

/transl_table=11

/product="3-hydroxyacyl-CoA dehydrogenase/enoyl-CoA hydratase"

/protein_id="BAB07207.1"

/db_xref="GI:10176112"

translation="MARQIRKVAVIGSGVMGSGIAAHLANVGIPISLLDIDVFNQLTTEE BORKGLTLEDAVRNRLATLALKLKOKPAPLSRKSHVHYIQPNLNEDDQAEKLAEVD WILVVENLDVKKQLEPKVDQYRKQGTLYSSNTSGISIAMKGRSDFQAHFLGTH FPNPRYLKLEIIPTRKTKEPVTAFMRAFAEDVLGKGVVEGKDTFNFNIRIGTYGL LVTQKMLEGYSVEYDSVTGPIGRPKSATFRTLVDVGLDTFLHVAKNHNVVDGE KQVDFDPVFMMKMAENWIGSKQGQFFLKQKGENGSEILELNPSTLEYEKRRKKKA PSIQAKQKSLNEKLSLIIQANDRAGOLLNMLKPLVIYSAEKYTAGDITKAVDDA MKGFWGBOGPFEMWDAIGVASSVORMEEGATVPWKEMLTSGRERFYODCTFYHH GEAKVIPNGKIHLKALKPERVITKNSCASLIDLDGDOVALLFHFSPNSIGFDVVOM INKAVDEYANYKGLVIGNQSKFCVGANLMLMEADDDNFEFDMVYRQFOQAMAK IRYSKPVSAPFAMTGGTEICLPTASIQASLETYMLGVLVEGVGVIPGGGNGKELY LRHLERPEGANIDQSTARKTTFETIAMKYVSTSAHEAMENGFLSERDGVVYNGDHL QYAKQQVILHHDAGYRPPARKKFPVVGTFYATMLLGAKSMKFGGMISEHDLKIAEKL AFVIAGGRVAKGTLVDEQYLLDLEREAFSLVGEPKSQORMOHLMLMKGLRN"

5746..6015

/gene="BH3489"

5746..6015

/gene="BH3489"

/note="BH3489

unknown"

/codon_start=1

/transl_table=11

/protein_id="BAB07208.1"

/db_xref="GI:10176113"

translation="MQQSFMQSGQQQGVMPQPOVITVKDQLYITDMLSWNLLACKK AHFTAQOCQDPDEVKQALEQAGQMQRHYTLLGLHQNQNMATVQ"

6030..6353

/gene="BH3490"

6030..6353

/gene="BH3490"

/note="BH3490

unknown conserved protein in B. subtilis"

/codon_start=1

/transl_table=11

/protein_id="BAB07209.1"

/db_xref="GI:10176114"

translation="MNOQKIQNPETPVQKNTQMSDRDFINDOLATEKYITSSSIAM NEASHQOLYQDIFTACQEAQDCQOLYELMFKKGWYLEAAPEQLQEAQQOFSGYM QLPHY"

complement(6396..6644)

/gene="BH3491"

complement(6396..6644)

/gene="BH3491"

/note="BH3491

unknown conserved protein in B. subtilis"

/codon_start=1

/transl_table=11

/protein_id="BAB07210.1"

/db_xref="GI:10176115"

translation="MDEKLQOLDGLVEKYAELLVGDTSSEMKERIRAYIYSHISK MPPLAKHWSQVPPAKEGMKALFLDIQLNKKQRENKQ"

6761..7150

/gene="BH3492"

6761..7150

/gene="BH3492"

/codon_start=1

/transl_table=11

/product="transcriptional regulator (GntR family)"

/protein_id="BAB07211.1"

/db_xref="GI:10176116"

translation="MNYTFDDSRPLFQQVADAIANDIHNVIKEGEQVPTNQFAKHH QINPATVAKGHLLDQGLFKKRGVGMFVAKGARDQLLEQRRETFYSDFIYPLLINEA KTLHLSSEIMLDLIKRGANDSDDH"

Query Match 9.7%; Score 91.4; DB 1; Length 303650;
Best Local Similarity 50.8%; Pred. No. 1e-05;

Matches 299; Conservative 0; Mismatches 281; Indels 9; Gaps 3;			
Qy 42	aaatatgtgagagattcagtagaatctatatataataacaaagccttactgatttgaatt 101		gene
Db 246029	AAATATTATTAAGCGAAGATAGAAAGTATCTTAACCAACAACTTTGAAAACCTTGAGTT 245970		CDS
Qy 102	cataatgttcattgataatccaagttaggagtgatttaagaacaattcttaacagataattc 161		
Db 245969	TCTAATAATTAAATGATGCTTCTACTGATGGAAGTGG---GGAATTGCTAGAGGAATACTC 245913		
Qy 162	agttgtagataagaataaaatcttgcttaataagaagaaaataattggttttagcatcaag 221		
Db 245912	AAAGAAGATAAAGCAATGAAGATTAAATTCACATAAAACAACTAGAGCTCAAGTTACAG 245853		
Qy 222	tttgaacaaagcggtagaaattcttaagggagaatatattttagaatggatgctgtatga 281		
Db 245852	TTTAGCTGAAGAGTGTCCTTAGCTAAAGCACCTTGGGATTCCAGAATGGATGCAGATGA 245793		
Qy 282	tattcatatccaagttagattgataagcaaatcgtttttttaggagaaaatttcattgga 341		
Db 245792	TGTTTCATTTAAAGACCGCTCTTGCTGTTCAATGGATCATGTTAAGGCCCACTCA--GAA 245735		
Qy 342	tttctcagcaactaataagatgatagaccaaaaggaataatttagtataataacaacg 401		
Db 245734	CTCGATACCTGTTAGTTATGTTATTGATATAGATGATAAAGGGAATGAGTTGGAATA 245675		
Qy 402	agaaagtaataaatactacttaactaatgatatacggaagatgttattgtaataagatctat 461		
Db 245674	AGAAGGTACCACTACACATAAG----GAGATTGCTAATTTAATATYGGACTTGCCCTTT 245619		
Qy 462	acttgcaccacccaaagcgtggtgcgtataaagaaagatttctgataagttaatggatatag 521		
Db 245618	TATTCAATCCGACAGCTGTGTTTAAAAAAGACTCAATTATCAAGCAGGATCATATGATCG 245559		
Qy 522	agatttagtaactgttaagaattatgatttgcataagaagagcctcgctgatttcaa 581		
Db 245558	TAATCTAAGACGGCGCTCAAGATTATGATCTATGTTTCGCTGCTTAGAAGCCAAACTCAA 245499		
Qy 582	aatcggttactcaataaagcactttacagtagatagataaacagagaat 630		
Db 245498	GTTTGAGNACATAGATAAACCATTACTCTATTATCTGTTCAACATGAT 245450		
RESULT 8			
PFMALIP3			
LOCUS	PFMALIP3 67970 bp DNA linear INV 15-DEC-1999		
DEFINITION	Plasmodium falciparum MALIP3, complete sequence.		
ACCESSION	AL031746		
VERSION	AL031746.9 GI:6594243		
KEYWORDS	HTG.		
SOURCE	malaria parasite P. falciparum.		
ORGANISM	Plasmodium falciparum		
REFERENCE 1	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
AUTHORS	Bowman, S., Churcher, C., Harris, B., Harris, D., Lawson, D., Quail, M. and Barrell, B.		
TITLE	Direct Submission		
JOURNAL	Submitted (24-SEP-1998) P. falciparum Genome Sequencing Consortium, The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK		
COMMENT	On Dec 16, 1999 this sequence version replaced gi:5763807. See http://www.sanger.ac.uk/projects/P.falciparum. IMPORTANT: This sequence is unfinished and does not necessarily represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the sequence may change as work continues. The sequence may be contaminated with foreign sequence from E.coli, yeast, vector, phage etc.		
FEATURES	Location/Qualifiers		
source	1. .67970		
	/organism="Plasmodium falciparum"		
	/strain="3D7"		
	/db_xref="taxon:5833"		
	/chromosome="1"		
	complement(join(1748..2598,2748..2848,2990..3276))		
	/gene="MALIP3.01"		
	complement(join(1748..2598,2748..2848,2990..3276))		
	/gene="MALIP3.01"		
	/note="MALIP3.01, conserved hypothetical protein, len: 412 aa, similarity: UPF0006 family eg to YBL055C/YBL0512/YBL0511, YBF5_YEAST (418 aa), fasta scores: opt: 316, E(): 1.1e-12, (33.2% identity in 271 aa overlap)"		
	/codon_start=1		
	/product="conserved hypothetical protein, UPF0006 family"		
	/protein_id="CAB63556.1"		
	/db_xref="GI:6594244"		
	/translation="MKLVFHYIKYINLVFYISIIIFLKSNSLKIYNDLRYISTVNVKYK LQIKRSNLKKNHNRKMDNESFIDIGSNLTDMFQDGYVNSKKHENDLQNLNRAK NNVDKIIITCTCLAEIDKSLKICETDYDEGKFLYLSAGVHPTNCYCFIDKNKHEE IIAKYEYEFIKYFKNEQVENSNGKNKKICDGEKDMNMLNEILLKKNLDITFGFKY NEKDKEYLENKKNRIKYPNRIKIGIGIGIDFLRYFCFSKYIQIKYFQKLKQMFN LPMLHMRNCSETPFKIVDIYKFLPEKNGGVHSFTDKEDIVHIVQNKLYIGVNG CSLKLEINAVKKIPINLLILLETDPACWCGVKKTHASVEYIKDYERRAYTNLAKIKN IIRCDNTIIFKERNEPYNIA"		
	complement(2599..2610)		
	/gene="MALIP3.01"		
	/note="potential splice acceptor sequence"		
	complement(2742..2747)		
	/gene="MALIP3.01"		
	/note="potential splice donor sequence, atg/gttaa"		
	complement(2849..2861)		
	/gene="MALIP3.01"		
	/note="potential splice acceptor sequence"		
	complement(2984..2989)		
	/gene="MALIP3.01"		
	/note="potential splice donor sequence, aaa/gtaaaa"		
	5005..5496		
	/gene="MALIP3.02"		
	5005..5496		
	/gene="MALIP3.02"		
	/note="MALIP3.02, hypothetical protein, len: 163 aa, contains possible signal sequence"		
	/codon_start=1		
	/product="hypothetical protein, MALIP3.02"		
	/protein_id="CAB63557.1"		
	/db_xref="GI:6594245"		
	/translation="MKLINRFVVLVLCPTIILFFFLNSVVLGNRRNNINPHETENAAK AMRKLISGEINSIKLDNGDELKIKLNDKHKDSTKWDKSYFISNLBEEKYSQDILFR KQKEINAEANTKIIDROEFYILNDEIENIATREFVLENNFDELYIQSFQSLIDLIQS LNN"		
	8020..10389		
	/note="possible cen1, region of very high [A+T] content"		
	14884..20352		
	/gene="MALIP3.03"		
	14884..20352		
	/gene="MALIP3.03"		
	/note="MALIP3.03, putative ABC transporter, len: 1822 aa"		
	/codon_start=1		
	/product="putative ABC transporter"		
	/protein_id="CAB63558.1"		
	/db_xref="GI:6594246"		
	/translation="MTTYKENVGISNKGKKNKKQCNISFLNLSFDWIRPLINDLIK GDIELNLCRNFDVPPYVASKLEENLRDIEVEDSEFYSEKNSNHEVLRHCHNSDASE KVVNVVYHNLTLSILKTFKRIITLISFYLETILVTLCGKFIDYMRILEGGOKIPV YISLDFKVPFSGLVVNMIMFFLFFALLHFFLHFTINTKLKSLMFLYKINLCSSNN NHLQNPDAFYNTYRKFSQTEIDSRDLSIGRNASDSSSGIKNNKNNIDNNKFVEN DYIINFIKSTKKMEKDSLENRSLPNVNIYMFSDVSVTFVTCINLNFVFKIF MSFYVFIKIGSVSGTAINLSIALYSAMILFPLPSFKSKYILTRYDRKIDNNHHVL KEFKLIKMFNWSFAFYINFRKMKYCKIRLYLSNIGVFISSIDIVDEVVFFI YLKDLANKKEIKFTSIIMPLYVVKILISNVANPNLVNVMVEGIVNKLKNLINDH LYNDIKNYPNRYRNYNDVVDKTFQENITSHDDGSHNLKHLKNKKLNTN MKYFFFTKMNHYHNKNIKOILSGLKNDVDDNTKKICFQEHKSNKNTYNTNSSHIH KKEENYHNHNSNWFNFKKKNNKNEYIILKLENCSEFLGSKDNKNDHLLKKNIN LKRNLSAIIIGNVSGSKSAFFHSILGDFNMTHGNLYIENFFKMPILYVFPQNSLWFMG		

Db 9146 TAATAATTTAAATAAATTAATAATTTAAATATATAATAATAATAATAACGCTAATAA 9205

QY 701 aattctattgatgatacacaaaataactactcttccaagatgtagtataaagaac 760

Db 9206 ATAAATTAATATATAATAATAATAATAATAATAATAATAATAATAATAATAAAT 9265

QY 761 gctactcgcgaagagctctctaaatttttgagctaaatctaccctcctagttacta 820

Db 9266 AATATAATAAATTAATTAATTAATAATAATAATAATAATAATAATAATAATAATAA 9325

QY 821 ttgaaactatattgtttattattattacatttaagctccctccttggtta 871

Db 9326 TAACATTATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 9376

RESULT 9

AC092304/C

LOCUS AC092304 Homo sapiens chromosome 19 clone LLNLFOS-20C5, WORKING DRAFT

DEFINITION SEQUENCE, 6 unordered pieces.

AC092304

AC092304.1 GI:14589493

VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.

KEYWORDS human.

SOURCE Homo sapiens

ORGANISM

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS 1 (bases 1 to 36977)

TITLE Sequencing of Human Chromosome 19

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 36977)

AUTHORS DOE Joint Genome Institute.

TITLE Direct Submission

JOURNAL Submitted (03-JUL-2001) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

COMMENT -----Genome Center Center: Joint Genome Institute Center Code: JGI Web site: <http://www.jgi.doe.gov> -----

Project Information Center Project Name: 31883, FOS36853 Center clone name: LLNL-FOS_20C5 -----

Summary Statistics Consensus quality: 31111 bases at least Q40 Consensus quality: 33034 bases at least Q30 Consensus quality: 33871 bases at least Q20 Estimated insert size: 42980; agarose-fp estimation Estimated insert size: 36477; sum-of-contigs estimation Quality coverage: 6.56 in Q20 bases; agarose-fp estimation Quality coverage: 7.73 in Q20 bases; sum-of-contigs estimation.

* NOTE: This is a 'working draft' sequence. It currently consists of 6 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence.

* as soon as it is available and the accession number will be preserved.

* 1 1580: contig of 1580 bp in length

* 1581 1680: gap of unknown length

* 1681 3748: contig of 2068 bp in length

* 3749 3848: gap of unknown length

* 3849 6070: contig of 2222 bp in length

* 6071 6170: gap of unknown length

* 6171 13635: contig of 7465 bp in length

* 13636 13735: gap of unknown length

* 13736 22951: contig of 9216 bp in length

* 22952 23052: gap of unknown length

* 23052 36977: contig of 13926 bp in length.

Location/Qualifiers

source 1. .36977 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="19" /clone="LLNLFOS-20C5" BASE COUNT 11122 a 7407 c 7254 g 10688 t 506 others ORIGIN

Query Match 8.4%; Score 78.6; DB 2; Length 36977; Best Local Similarity 46.7%; Pred. No. 0.0015; Matches 433; Conservative 0; Mismatches 475; Indels 20; Gaps 5;

QY 27 atataagagcctttaaattatgtagagattcagtagaactctataataatcaaacgct 86

Db 34445 ATATATAATTTATATATAATAATAATAACATATATATATATATATATATAATA 34386

QY 87 tactgatttgagtcbaattgctcattgataatcccaagtagaggtgatttaagcaatt 146

Db 34385 CATTATATATAATTTATATATAATAATAACATATATATATATATATATATAATA 34326

QY 147 cttaacagaatattcagttgtagataatagaataaaaaatcttgcttaa-----tgaaga 200

Db 34325 AATACATTTATATAATTTATACATAAATTTATAAATACATTTATGTATATATACAATTTATATA 34266

QY 201 aaatattggttttagcatcaagtttgaacaaagcgtgaaaatttcttaaggaggagaat-ata 259

Db 34265 CTATATAATTTATATATATACTATAATAATATTTCTTATATAAATTTATATATAATAATTT 34206

QY 260 ttttgaagaaggatgctgatgataatttcoataatcccaagtagatttgataagcaaatcglt 319

Db 34205 TATATGTTTATATAAATTTAT 34146

QY 320 ttatggaggaaaatctcattggtatttcagcaactcctaataagaa-----ttgataagcc 373

Db 34145 ATATATAATAAATTTATATGTTTATATAAATTTATATATATATAATAATAAATTTATATATATAT 34086

QY 374 aaaaagaaaatttagtataataaacaacgagaaagtaataaataataataataataataataata 433

Db 34085 TATAATTTATTTATATATAATAATAATTTATATGTTTATATATAAATTTATATATATATAATAAT 34026

QY 434 tacggaagatgtatttgaat----agatctatactctgcccccccaacggtgtagtaaa 489

Db 34025 ATATTTTATATGTTTATATAAATTTATATATATAATAATAATAAATTTATATATATATATATATAT 33966

QY 490 aagaagttttcgtataagtttaattgggataagagatttagtacctgttgagattatgat 549

Db 33965 ATTATATATAATAATAATAAATTTATATGTTTATATAAATTTATATATATAATAATAATAAATTTTA 33906

QY 550 ttgcaataaagaggagctcctggtgatttcaaaatcggtctactcaataaaagttactcttta 609

Db 33905 TATATTATATACTTACATATATATATATAAATTAATAATAAATTTATATATATATATATATATATAT 33846

QY 610 cagtatagattaaacgagaaatggaatatcaacaacccaataagtttaagcaataatatttacc 669

Db 33845 TATTATATATATAAATTTATATGTTTATATCATATTTTATAAATTTATATAATAATAATAATATG 33786

QY 670 tcagctattttacaagatttttataaagaaaatcttatattgatat---cacaaaaatt 726

Db 33785 TATAATATATACTTATATAGGTTTATGTTATATAATAATAAATTTATATATATATATATATATATAT 33726

QY 727 actaattactttcaagagtagtattgataaagaagcgtactactcagcagagagctctctctaaa 786

Db 33725 ATATA 33666

QY 787 tattttgagctaaatctaccctcctagtagtattactattgagaacatacatatttggtttcat 846

Db 33665 TATATAATTTATACATACTGTTATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 33606

QY 847 ttatacttttaagctcccttggttaggaggtttataataaataaatgataattattttta 906

Db 33605 TAATATACTATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 33546


```
VFSSMLKFSKKYNQLDFILYYIGVSFSYFFKKPSYALLRLRNMFLRNK"
gene      6944..8191
/ gene="wzx"
CDS       6944..8191
/ gene="wzx"
/ note="wzx"
/ codon_start=1
/ transl_table=11
/ product="O-antigen flippase"
/ protein_id="AAL27353.1"
/ db_xref="GI:16611772"
/ translation="MNFVLQILYLILANSRRIPIPLPLSARLLNSDPEALSTII
SFATWAIVIEFGFNYSKQIKIYDSNKVKVISVINSTLSAKILSLAILGIYTSF
YVGTQHFYIVYLLWLSYFFVGMLSFYVICQNNKQLLKVELIGGVAFLMIFVWFS
NKNFYFIVILVAYRCMVELLSIFFLKNRVDLSKIRLKSSGINAKKTYSYALFOI
SSSILYGLSIAICVIDKNIILYHLLAERVEKLASFPSQATRVIVYSELNFTGNKK
NIIWVPCSGVACLGIIIGLFIKTLFIVIEIYFGKEFIAAHLNLDILALPPAPFLN
GIMSVGFLTKGNIKYLNVTIIITAGCVSIPVCYFLSSSYPOISTSIAYVVAEGLISLI
LLIKFVELLKENKYESREDNIK"
gene      8163..8696
/ gene="rmlC"
CDS       8163..8696
/ gene="rmlC"
/ note="rmlC"
/ codon_start=1
/ transl_table=11
/ product="dtp-6-deoxy-D-glucose-3,5 epimerase"
/ protein_id="AAL27354.1"
/ db_xref="GI:16611773"
/ translation="MKVEKTLINDAFIEIPDVYDGERFFQEVYHHBKYNIIENGT
FVQNDYRSRQKNVLRLGHFQTSNPQGLVRVRSQVDFDAVDLRKNSSTYGHWYGIEL
NEWNNKQFIPPMQMAHGFLVLSOIADPEYKCTDFYHPEFEKLIWNDPINIWNKPINE
EPILSPKDRKGRFEDL"
gene      8698..9471
/ gene="wbgS"
CDS       8698..9471
/ gene="wbgS"
/ note="wbgS"
/ codon_start=1
/ transl_table=11
/ product="putative glycosyl transferase"
/ protein_id="AAL27355.1"
/ db_xref="GI:16611774"
/ translation="MDELKPLVSIIMPAYNSQYTIKDTIQSVLKQTPYNNWELLITDD
LSSDSYPKIKKAAOPRIKLYSNPNVGAAYSRNNSLKNANGCETAFDLSDDLWIA
SKLEKQILWMLKPNITFSTAYQLIDEHGNKLKLIIDFOGDNVSCFYHMLPEKATL
GCSTVILKDAFDIDKMPLIIRTQGDYALWELKLSKERCYLYNDLVLTQYRVMPNSISR
NKTKAKRQWEIIRKIEGLPLVYSVICFISYIGWRAIFRR"
gene      9595..>10911
/ gene="gnd"
CDS       9595..>10911
/ gene="gnd"
/ note="gnd"
/ codon_start=1
/ transl_table=11
/ product="6-phosphogluconate dehydrogenase"
/ protein_id="AAL27356.1"
/ db_xref="GI:16611775"
/ translation="MSKQQLGVVGMVMBGRNALNIESRGYTVSIFNRSREKTEEVIA
ENPKGLIPIYTVKEFVESLETPRILLMYKAGAGTDAAIDSLKPYLDKGDIIIDGN
TTFQDITRRNLSAEGFNFIGTVSGGEGCALKGPSIMPGQKEAYELVAPILTKIA
```

```
Query Match      8.3% Score 78 DB 1 Length 10911;
Best Local Similarity 49.0% Pred. No. 0.0025;
Matches 328; Conservative 0; Mismatches 330; Indels 12; Gaps 4;

Qy 13 atcattatgctggtatataatgagocctttaattatgtgagagatcgatgagaattctata 72
Db 8728 ATTATTATGCGCTATAATTCGAGT---ATACATAAAAGATACATACAACTCTGTT 8784

Qy 73 ttaatacaaacgctactgatttgagttcattgattgataatcccaagtagaggt 132
Db 8785 TTTAAACAAACATATCCAAATTTGGGAACATATTGATT---ACAGATGATCTCTCTCTGAT 8841
```

```
Qy 133 gatttaaaagcaattcttaacagaataattcagttgtagataataagaataaaaaattctgctt 192
Db 8842 TCACACCGCAAAATTTATAAGAAATTTGCTGCTCAAGATCCTCGGATAAAATTTGATTCT 8901

Qy 193 aatgaagaaaaatattggtttagcatcaagtttgaaacaagcgggtgaaaaattcttcaaggga 252
Db 8902 AACCCAGTAAATTCGTGGTGGCGGACGTTTCCAGAAAATAATTCACATAAAATGCCAATGGG 8961

Qy 253 gaatatatttttagaatgagctgcatgatatatttcatatccaagttagatttgatagaagaa 312
Db 8962 GAATTTATTGCTTTTCCTTGATAGCCGATGATTTGTGGGAGGCATCTAAACCTTGAAAAACAA 9021

Qy 313 attcgttttttagggagaaaaattca---ttggattttctcagcaactctaataagaattgata 369
Db 9022 ATATTATGATGCTCGAAAAATCCAAATATAACATTCTCATTTTACAGCATATCAACTATT 9081

Qy 370 gaccaaaaagaaaaatttagttatataaacaacgagaaagtaataaaataacttaactaat 429
Db 9082 GATGAGCATGGAATAATAATTAATAAAAAAATTATTGATTTCCAAAGGGGATAAATGTTTCATTT 9141

Qy 430 gatatacgggaagatgttattgaatagatctatactgccccaccacaacgtyggcgtaaaa 489
Db 9142 GGTATCATGATGATGCTTTTAAAAAAGCAACCTTAGGTTCCTCCACAGTAATAATTTTGA 9201

Qy 490 aagaaagttttcgataaagttaatggatagagatttagtacctgttgaaagatttatgat 549
Db 9202 AAAGATGCTATTGTGATCATATAA---AAATGCCATTAAATAGGACATGGTCAAGATTATGGG 9258

Qy 550 ttgcaataaagaggagcctcgtgatttcaaaatcgccttactactcaataaaagtactttta 609
Db 9259 CTGTGGTTAAAAATTACTAAAACTCTGTGAACGTTGTTATTATACAAATGATGTTTGTGACT 9318

Qy 610 cagtatagattaaacgagaatggaatatcacaaacaaataaagtttaagcaaatattttac 669
Db 9319 CAATATAGAGTTATGCCCAATTTCTATATCTAGAAAACAAAATAAAAGGCAAAACGTCAA 9378

Qy 670 tcagctattt 679
Db 9379 TGGGAGATAT 9388

RESULT 11
VCLPSS
LOCUS      VCLPSS              16265 bp      DNA      linear      BCT 30-NOV-1997
DEFINITION V.cholerae ORF's involved in lipopolysaccharide synthese.
ACCESSION  Y07786 Y07787
VERSION     Y07786.1 GI:2244680
KEYWORDS    galactosyltransferase; mannose-1-phosphate guanylyl transferase;
ORF21x9 gene; ORF34x7 gene; ORF35x9 gene; ORF39x2 gene; ORF40x0
gene; ORF40x1 gene; ORF41x3 gene; ORF41x8 gene; ORF41x9 gene;
ORF50x8 gene; ORF54x5 gene; ORF56x5 gene; oxido-reductase; Rfc-like
protein; sugar transferase.
SOURCE      Vibrio cholerae.
ORGANISM    Vibrio cholerae
REFERENCE   1 (bases 1 to 16265)
AUTHORS     Manning,P.A.
TITLE       Direct Submission
JOURNAL
REMARK      Revised by submitter 20-OCT-1996
REFERENCE   2 (bases 1 to 16265)
AUTHORS     Stroehner,O.H., Farasivam,G., Dredge,B.K. and Manning,P.A.
TITLE       Novel Vibrio cholerae O139 genes involved in lipopolysaccharide
biosynthesis
JOURNAL     J. Bacteriol. 179 (8), 2740-2747 (1997).
MEDLINE     97252505
FEATURES    Location/Qualifiers
             source
               1..16265
               /organism="Vibrio cholerae"
               /strain="AI-1837"
```



```
gene
CDS
/sub_species="O139 Bengal"
/db_xref="taxon:666"
190..1308
/gene="ORF41x9"
190..1308
/gene="ORF41x9"
/feature="putative"
/codon_start=1
/transl_table=11
/product="oxidoreductase"
/protein_id="CAA69111.1"
/db_xref="GI:2244681"
/db_xref="SWISS-PROT:Q56598"
/transl_table="MKKALITGVTDGQSYLAPELLAKGYEVHGIKRRASSFNTOFV
DHIQDPHVDNASFLHYGDLTDSNLTIRILQEVKPDVYNLGAQSHVAVFSPESPEYT
ADVDMGTLRLLEALRLGLEKTTKFCOASTSELYGLVQETPOKETTFYPRSPYAVA
KMYAVIIVVYRESGYACNGILFNHESPRGETFVTRKITRGLANIAQGLEKCLYM
GNMDALRWGHAKDYVRQWMLQODQDFVIATGVQISVRQFIEWSAKELGVLTFT
EGQGVDEKIVTATEGDKAPALKVGDVVVQIDPRYFRPAVEETLLGDSPKAKQLGWT
PEITVQENKCAEMVMDLKVQARHALLKLHGHDPVPSVE"
1313..2239
/gene="ORF34x6"
1313..2239
/gene="ORF34x6"
/codon_start=1
/transl_table=11
/protein_id="CAA69112.1"
/db_xref="GI:2244682"
/db_xref="SPTREMBL:O34227"
/transl_table="MKILLTGARGVQGNILSLAGEYEHFELSPSSAQLNLLDSNAVH
DYIRHKPDWPHIAAGIVGGIOANMAQPVKFLVENMQMGNILTVSNACGVSQELNS
SSCYTPDRAHNPBELLKGELEPTNEGVALKLEITSTRLCYEYISKENPSOIKYTIIP
CNLYGRFPDQPNHSHMIPAVIRKTHEAKANLEPIDTWGGEARREPMYAEADAEFV
FYADHKEMQNLNVLGTDYITNEYQAVASVIGYTGGRKPHDLSKVPQGNRKLIDD
TKLAFGSHSTSLSEAGIEKTYQFYLYNEIGND"
2232..3404
/gene="ORF43x9"
2232..3404
/gene="ORF43x9"
/codon_start=1
/transl_table=11
/protein_id="CAA69113.1"
/db_xref="GI:2244683"
/db_xref="SPTREMBL:O34228"
/transl_table="MINYSLASTWQDQNEAIAQRIQVIESDMFTMGKEVAQYKDFATF
FGAVALAVSSGSTANLLMIAALFTKNAKYKLERGDEVIVPAVSWSTTYPLQOYGL
KVKFVIDROTLNMDLKLKLEAIDTNTRAILSVNLGNPNDFARMKDIIADRDIFILE
DNCSMGATINGKQAGTFGVMTGTFSSFSFHIAHTMEGGCVYTDDEELKHILLICIRAHG
WTRNLPRFNKVTGKSDDAFEESEKFLVLPNGYNVRPLEMSGALGIEQLKPLNFIETRR
KNAELFQSLADHPFIEIOKETGESSWFGSLILKENAPYSRAELVKKLVNGIECRP
IWTGNFLNKEVLEFFDYEVAGTLEGAEYLDENGLFVGNHQNDIEAEIKLLAKLVV"
3404..3889
/gene="ORF18x8"
3404..3889
/gene="ORF18x8"
/codon_start=1
/transl_table=11
/protein_id="CAA69114.1"
/db_xref="GI:2244684"
/db_xref="SPTREMBL:O34229"
/transl_table="NMLEKDFRGLVQHGVLFSDLVVLTQNTQLGVRVNRPAQGY
WFVPGGRVYKNETLIKAFERICDELGOIFSQDAKPLGLYHFNDSVFGDDISTHY
INAPHLIESENTTLNLPSEQHRHYRWVIRIGDLAYDESVEHFSYKVFLELKKRIAFTI
I"
3946..5382
/gene="ORF54x4"
3946..5382
/gene="ORF54x4"
/codon_start=1
/transl_table=11
/protein_id="CAA69115.1"
/db_xref="GI:2244685"
/db_xref="SPTREMBL:O34230"

```

```
gene
CDS
/translation="MIINAFYRDFLLSILGKVLQVFIGLTRLITELSEQVGIIYI
LLTVSLAFGFENPLQGYGRYLTHMOKAELKTATNMLMLRGLTIPFALCLAIIV
FYMRPEOYFSIAEYSVFIVSVIALIHGVLNATNVLVSRVKFTYAVLTLFGLIL
SVYFQKOTAMANIYGLAFSQIFSFVILNITSEQEFIVKISAKNHYMKGVLP
FTLPVITLLQWQOTAFRLIVEDLYSAQILAFIAGVMTLGSALFSALESLSQFYM
PLYLKKITNADOAMRTKAWNELADILLEVYIAVLIYIATAPYKAKLLVAEKFOQAYI
YAMIGAFIEFFRVITNLVYVQSSEVKTSTIMPLYLIGFMMVMVIGLYIDVSRSPWL
PVILALSYLTLSIMFYNMKLLSIRFEFLNPAIPATAISLTLPLLVHMLLSITPTLNSF
LSLLGGFYLLLIQYPELTKIKKMSAV"
5379..6395
/gene="ORF40x0"
5379..6395
/gene="ORF40x0"
/codon_start=1
/transl_table=11
/protein_id="CAA69116.1"
/db_xref="GI:2244686"
/db_xref="SPTREMBL:O34231"
/translation="MKKACLVVNDYMLCNKVFDEKIFRDNINIKYIKLKLERYGYD
LSTODINSISDVVYIADIPKPLKRNKVKSHLIIWSEFISPENYKVKHNSFN
KIFTWHDGLVDEKKYKFINVQHPNFINKNKKKLTCTLIAGNKKPHTSVGLYA
ERKALRMFKENHPSDFDLYGVGNDRYFSGSKIYRALNRVPLLPKTKOKLTGRSYPS
YKGPIERKRPIMEKYKFSICYENVMDFCYITDKLFDSFFSGCIPVYWGANNIETIIP
ENCFDIKRKFLSYEDLYSYMKEMSDDEEYIGYINNIQAFILKSDLSPFFKAETFAEKVVS
EVLN"
6395..7408
/gene="ORF40x1"
6395..7408
/gene="ORF40x1"
/codon_start=1
/transl_table=11
/protein_id="CAA69117.1"
/db_xref="GI:2244687"
/db_xref="SPTREMBL:O34232"
/translation="MKIRLIVAFKIEYGRKENGFPALYKAWFHGFEQJGYSIEGVFF
EDYDHTMLOQRIIEKAMNSIDIVFIILOQDQVDLTLYLSENNVFTVFGFDLWR
FDNWTSIYAPFSACITSDKFSIKKYEIGQGNIIYSQWASLEPKQYHSAPYKYDISF
IGSANMYRKWFINLEKKYIHVECFGYGKNGRVSYEDMELIFLQSKYNLINSISIF
DFRPMSSFRSFLSVKSLYKVDGKNSQTRARNEIPIVHGGFOITDVPPTLEEVFKI
GKEIVCYNTIDEVEKLIRFYLINDYERECIKRACHTYKARIKNFMVNDIKI
YDK"
7398..8855
/gene="ORF56x5"
7398..8855
/gene="ORF56x5"
/codon_start=1
/transl_table=11
/protein_id="CAA69118.1"
/db_xref="GI:2244688"
/db_xref="SPTREMBL:O34233"
/translation="MISIKIGISKLIEPHFFNELEPENYKVLTCSRSRLTWTFRDLAF
KLLYLEMDKNVSFSKEAVKEHIRAFSLGPGKEQSKNSIEKPYDAFFETFNDSL
HGFDATKSLIPLSHNGSTANGAHRVASAIILDKDVCYKLPVCDHLYDKFPYSRSVS
CDLDDIAATKFEYADNVIYIAFVWPTAGCFDEETRIIPNIIYKRNIRMTPNGANLL
SOIYFGEPLWGTAEENNFRGSKNTECFKTFDMRVIAFRADSLDSVLQIKENIQOIF
NVGKHSIHITDTKGEAIRMARMIENDSIHFLNAYPNKYSTHEKLEAFKHHIDNC
IGSDDIILDSGMVLSIYGLREASDIDYLSIKSLEYKNEGLECHEKLEYHDEKNEL
IYNKYPPFYENGLKFIATNOLYKMSNRDEKDRNDCKMESLIENNQYKNIKAKLKQ
SIYVEKIKLRKIKTCLLASIGLYDLVKKIYKVLK"
8852..9865
/gene="ORF39x2"
8852..9865
/gene="ORF39x2"
/codon_start=1
/transl_table=11
/product="sugar transferase"
/protein_id="CAA69119.1"
/db_xref="GI:2244689"
/db_xref="SPTREMBL:O34234"
/translation="MSSPKISVVMVNGEYKVLGAIDSILKQTFSDFEFIINDGST
DKTLFRIKSNMKDDRIVLSPENKGLIVSLNEGDLAKGOYIARMDDADDSIKSRFE
KQIEFLDNDPDIQVCGTWEVFEVGENIKSKKKMKMPQDDDLKAKLTFSPFAHTVHIR
KKVIDKYKINDYDLGYKDAEDYKFWVDYSKYTLFSNVPEILLIRYHQESISKVAONKE

```


NKRRFRIISKIONEVLTSGVIVLTNEGERKHHFILTSLNERIINNVTDCDMIRALLKIKI SSQIESQDSSAIERMLMKYFIYLIILSIRRDKLSYLKIFDLMFLKGAFLFKDRM EQP"		10797..11861		/gene="ORF41x8"		
gene	9868..10800	/gene="ORF35x9"		10797..11861		
CDS	9868..10800	/gene="ORF35x9"		/codon_start=1		
	/transl_table=11		/protein_id="CAA69120.1"		/db_xref="GI:2244690"	
	/db_xref="SPTREMBL:P74948"		/translation="MTFLDISVVIPLNKNYSIVRCLNSVLYQITLPEFIIVNDGSTD DSLVLDYILKTIISMDRVNVVHQQNQGVSRANNIGIALAKSNYIALLDADDEWHPE HIGKMTLLTEKYPRLPVYCKHVCENGYSFTPOIFSSLETTMGLIDNYFVRAGKVE LVNSSKVLVYKPIFESVGGFPIGAKLCEOLYLWARLAEVGCCFAFTYLGVTVOEKDM SRKARNYSQPIILEYFYNVSGSINSELNHYLWTVYRNHLRLSLNLSLSEFISRWKFG SYFKPSKLIIFLYALIPKKIMNIFKVRIRFLR"		10797..11861	
gene	10797..11861	/gene="ORF41x8"		10797..11861		
CDS	10797..11861	/gene="ORF41x8"		/codon_start=1		
	/transl_table=11		/product="Rfc-like protein"		8.2%; Score 77.2; DB 1; Length 16265;	
Query Match		Best Local Similarity		Matches 183; Conservative		
0y		17		Db		
8877		TTATGTCTGTTTATAATG---GTGAAAAATATCTTGGAGAAGCCATAGATAGTATCTTA		8933		
0y		77		Db		
8934		AGCAACATTTTCTGATTTTGAATTTTATCATTAATGACGGTTCGAC-----TGATA		8987		
0y		137		Db		
8988		AGACACTAGAGATAAATAAGTCATATATGAAGAAGGATGACCGTATTGTTATGTTTCAA		9047		
0y		197		Db		
9048		GAGAAACAAGGCGTAATAGTAAGCCTAAATGAAGCGCTAGATTTAGCTAAAGGCGCAAT		9107		
0y		257		Db		
9108		ACATTGCTAGATGGATGGACAGATGATTTTCAATAAAGTCACGTTTTTGAAGAAGCAAAATG		9167		
0y		317		Db		
9168		AGTTTATAGATAGTAATCCA		9187		
RESULT 12		27730 bp		DNA linear		
AF461121		Escherichia coli		serotype O55:H7		
LOCUS		27730 bp		DNA linear		
DEFINITION		Escherichia coli		serotype O55:H7		
ACCESSION		AF461121		GI:18266395		
VERSION		AF461121.1		GI:18266395		
KEYWORDS		Escherichia coli.		Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;		
SOURCE		Escherichia coli.		Escherichia.		
ORGANISM		1 (bases 1 to 27730)		Wang, L., Huskic, S., Cisterne, A., Rothermund, D. and Reeves, P.R.		
REFERENCE		The O-antigen gene cluster of Escherichia coli O55:H7 and		Identification of a new UDP-GlcNAc epimerase gene		
AUTHORS		Unpublished		2 (bases 1 to 27730)		
TITLE		2 (bases 1 to 27730)		2 (bases 1 to 27730)		
JOURNAL		2 (bases 1 to 27730)		2 (bases 1 to 27730)		
REFERENCE		2 (bases 1 to 27730)		2 (bases 1 to 27730)		

Wang, L., Huskic, S., Cisterne, A., Rothermund, D. and Reeves, P.R.		Direct Submission		Submitted (18-DEC-2001)	
JOURNAL		Sydney, NSW 2006, Australia		The University of Sydney,	
FEATURES		Location/Qualifiers		1. 27730	
source		/organism="Escherichia coli"		/serotype="O55:H7"	
gene		/db_xref="taxon:562"		204..1199	
CDS		/gene="gnc"		204..1199	
		/gene="gnc"		/codon_start=1	
		/transl_table=11		/product="UDP-GlcNAc 4-epimerase Gne"	
		/protein_id="AAL67550.1"		/db_xref="GI:18266396"	
		/translation="MNDNVLLIGASGFGVTRLLLETAIDAFDNKLNLDKQSHFYFPEITQ IGDVRDQALDQALDAGDPTVLLAAEHRODVSPTSLSYIDNVNOCVTRNLAAEMKNGYK NIIFTSSVAVYGLNKNPDENHPDHPENHYGKSKQAEVLYREMYKAPTERSLTIIR PTVIFGERNENYVNLKQJAGCKEMVAGTANYKSNAYVGNIVFEFKYKLNKNAAGY EYNYVDKPLDNNQLVAEVEQSLNKIPSMHLLPYPLGMLGCGYCFDLSKITGKKYAV SSVRVKFCATTQFDATKVHSSGFGVADPTLTSQGLDRTLQYEFVHAKKDDITTFVSE"		1442..2335	
gene		/gene="galf"		1442..2335	
CDS		/gene="galf"		/codon_start=1	
		/transl_table=11		/product="UDP-glucose pyrophosphorylase Galf"	
		/protein_id="AAL67551.1"		/db_xref="GI:18266397"	
		/translation="MTNLKAVIPVAGLGMHMLPATKAIKPEMLPIVDKPMIOYIVDEI VAGIKELIVTHASKNAVENHEDTSYELESLEQVRKQRLLEAEVQSPGVTIMNV RQEPGLGHSILCARPAIGNPVPVYLDVDDASADPLRVNLAAMIARENETGRS QVLAIRMPGDLSEYSVIOTKEPLDREGKVSRIVEFIEKPDQPTQLDSDIMAVGRYLS ADTWPELERTOPGAWGRIQLDTAELAKKOSVDAMLMTGDSYDCGKKMGVMAQAFVY GLNLREGAKFLKNIQLAK"		2716..3810	
gene		/gene="wbqM"		2716..3810	
CDS		/gene="wbqM"		/codon_start=1	
		/transl_table=11		/product="putative galactosyltransferase WbgM"	
		/protein_id="AAL67552.1"		/db_xref="GI:18266398"	
		/translation="MVKILHVHLSGDISGQVRSVLDHMTLSQEFQOQSMVCSKBGRLA EQARFCVCVTHIPTLTREISLFLKDCASLFLYKIKKKEFDIVHTSSGTFLGRVA AKLAGTKKIVHTVHGFAPFPSTENKLIKFVFLMELIASYCSNIIIVNESDERTARKY FVKKSKLLILINNAIDVKYNDKDKDKDIPKIVMGRLCDOKNPLLLIPEAKDL ESNIHVDIGDGPLKVKLEKINQYNIADKVSFLGWDVAVEEHLKYKDLFVPLSRWEG MPLAMLEMAAKVPVLSSDIEANKYLIEKTAGVVVFKDEDSKDLKRNVLHANPELRN NLAKKAYQALIEDFDLTKRTKILLESYLN"		3831..4949	
gene		/gene="gmd"		3831..4949	
CDS		/gene="gmd"		/codon_start=1	
		/transl_table=11		/product="GDP-mannose dehydratase Gmd"	
		/protein_id="AAL67553.1"		/db_xref="GI:18266399"	
		/translation="MTKVALITGVTDGQSGYLAEFLLEKGYEYHGIKRRASSFNTERV DHIYQDPSAKPNPHLYGDLTDTNITRILQEVKPDVYNLGMASHVAVSFSEPT ADVDATGLLEAIRLIGLEKTRFYQASTSELYGLVQIPQKETTFFPFRSPYAVA KLWAYITVNYRESYGIYACNLGFNHPSPRGETFVTRKTRIAIANIAGLESCLYL GNMDSLRDGHAKDYVRMOMMLQEOPEDFVATQVQSVREPEVMTAOLGKLSF EGKTVDEKGIYVSVTGKAPGCVGDIIVADPRYFRPAEVEVETLLGDPKAKHEKLGHK PEITLREMISEMIVANDLQAKKHLVLLASHGFGNANLAQE"		5588..7045	
gene		/gene="manC"		5588..7045	

[illegible]

```
/note="probable transport of capsular precursors"
/codon_start=1
/transl_table=11
/protein_id="BAA33592.1"
/db_xref="GI:3724303"
/translation="MTQGYILRTLSLVACLSPVMAAQTPTPEQIQMFKNLPAAQQQA
LASKYFGDLPVSNASQYDNPVLMPPROPENITAPPLNGQDAPLERFGLSLFSG
SPSTFAPVDVPSDYTVGAGDEIVIQLFQKNDQTHRLRNVRAQVINFPSLGPVOVA
GLKSEVRSQSGQVEQIMIGIRLSIGELTMQIFVGMGAYKPGATVSALTLSQ
AIYSGSFQSQALRNIOIKREGVLVOLDLYDLKGDNRNDVRLPGDVVFICAVD
NVTVEGVNRPVYEVKSTETQDIIKMGAGFTANAYRDKTEVKRYASDGIKRALTL
NLANQGLQTLKLNNGDALNVLKKTQALAHYVEIGDVGPFGFIEWKPKLRVADVFSQV
DSAFNADLVNVLVRQKNLQDLEVOQIINLANAILEPFGSRDNLTLQSRDRLLIFNR
FNYYDFNDTSDKTEENTAKTLEQAQTAQKLEQEQKQOQINNTVRVYNASALANPS
LPRNNSQLPLSGQSLEAHPEAKIQFAGQEMTLAEFEKLNKONTRNLLAPILLQLYQ
QAQVMSQPIAEVGEVKYPGRYPIITAAMPYSALLEAAAGLTFNAYTLRAELAREID
PNHERVTQIISLDRMLSEPSANRWITOGMDKLNLEKPAKPAKLOQTVVLOGEVLFP
GTYTVROQETLAEKLRAGGLTFANPRGAITREALRLQEQKLNQYADMDRAETAK
KTFRADTNLSSVSDPAKTLAFVEASRSKALGRMVQLSRIMRDEPDAEMLEGDGF
LFPVTFRTISIMGEVQPVTVLLDGLNVDYLNKAGVKKQADADRIFVVRADGVS
YKPNSGYVWFGNKKELQAGDTIVPDIYDRDALSTWTAAQTQILYGTGVAINALK"
/gene="wzz"
/genes="wzz"
18485..19492
18485..19492
/gene="wzz"
/note="probable regulation of O-antigen length"
/codon_start=1
/transl_table=11
/protein_id="BAA33593.1"
/db_xref="GI:3724304"
/translation="MQEOKLPNPVLTYPQAQTAIGDDEIDLRELFLRMWKGKWTIIA
ITLFAIGSVIFAINOPNIYKAEALIAIPASEQQGGLSALASQFGLASLAVNGLGK
WVREKAPKEPSPMOEAYKVFSLVTKNADSGMWTIAVEHLSPTVAQOAVTWLIE
DINKVMKRDVAEHRSTAFLEQIALTNVADIRTVLKLIEQAKTMEAFVRDEYV
PKTIDPALAPEKAPKPRALICVLGTLGGMLGVAIVLRFAPKAGKNNHEDVRET
A"
19650..20777
/gene="IS1358"
19650..20777
/gene="IS1358"
/note="probable insertion sequence like element"
/codon_start=1
/transl_table=11
/protein_id="BAA33594.1"
/db_xref="GI:3724305"
/translation="MSELNPEMFQIIKDYROESKVEHKLSDIILLTCGVLSCHDG
WDGIIDFGNRLDLKRYGHEAGIPSDTLRSVMGMINPVALORSFIAWKNCHTLIT
DGEVTAIDGNLRLGKSDRSKGTIHMVNAFATANGMSIQGLKVDKSKNEITAPKLL
DLDDVKGCLITIDAMGCKKIAQKIRKREADYLLAVKGNQMGLEQAFDDYFRMDMLQD
FDGSSYTSQEKSHGRIETRVALNDRDLVLGDIHEHWPGLKSMGIVASIROESAVATE
QDVSIYVICSKELEAQTLLLEATRSHGVMEVHMWSLDTAFCEDNSRIRADDRAEAFAR
IROICNLKSETTEPKGIIKRRKMCAMDEKYLKSLVLESLT"
20961..22079
/gene="gmd"
20961..22079
/gene="gmd"
/note="probable GDP-D-mannose dehydratase"
/codon_start=1
Query Match 8.2%; Score 77.2; DB 1; Length 46721;
Best Local Similarity 57.2%; Pred. No. 0.0023;
Matches 183; Conservative 0; Mismatches 128; Indels 9; Gaps 2;
QY 17 ttatgtcgtatataatagagcctttaaattatgtgagagattcagtagaatctattaa 76
Db 29651 TTATGTCCTTTTATAATG---GTGAAAAATATCTTGGAAGACGCATAGATATTCTTA 29707
QY 77 atcaaacgcttacttttgattgattcattgtcattgataatccaagtagagggtatt 136
Db 29708 AGCAAAACATTTTCTGATTTTGAATTTATTATCATTAATGACGGTTCAC-----TGATA 29761
QY 137 taaagcaattcttaacagaatttcagttgtgagataatagaataaaaaacttgcttaag 196
```

```
Db 23762 AGACACTAGAGATAATAAAGTCTATATGAAGAAGGATGACCGTATTGTATTGTTTCAA 29821
QY 197 agaaaaatttggtttagcatcaagtttgaacaaagcggtgaaatttctaaaggagaat 256
Db 29822 GAGAAACAAAGGCGTAATAGTAAGCCCTAANTGAAGGCGTAGATTTAGCTAAAGGGCAAT 29881
QY 257 atattttagaatggtgctgatgatttcattcatcatccaagtagatttgataaagcaaatc 316
Db 29882 ACATTCTAGATGGATGCAGATGATATTTCATAAAGTCACGCTTTTGAAGACCAATTG 29941
QY 317 gttttatgagggaaaaattca 336
Db 29942 AGTTTTAGATAGTAATCCA 29961
RESULT 14
AC106820/c
LOCUS
DEFINITION
Homo sapiens chromosome 16 clone RP11-715J22, WORKING DRAFT
ACCESSION
AC106820
VERSION
AC106820.1 GI:18139370
KEYWORDS
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 177509)
AUTHORS
DOE Joint Genome Institute.
TITLE
Sequencing of Human Chromosome 16
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 177509)
AUTHORS
DOE Joint Genome Institute.
TITLE
Direct Submission
JOURNAL
Submitted (12-JAN-2002) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 1569302
Center clone name: RPCI-11_715J22
-----
Summary Statistics
Consensus quality: 173535 bases at least Q40
Consensus quality: 175401 bases at least Q30
Consensus quality: 175941 bases at least Q20
Estimated insert size: 176000; agarose-fp estimation
Estimated insert size: 177009; sum-of-contigs estimation
Quality coverage: 11.95 in Q20 bases; agarose-fp estimation
Quality coverage: 11.88 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
* 1 1009: contig of 1009 bp in length
* 1010 1109: gap of unknown length
* 1110 11850: contig of 10741 bp in length
* 11851 11950: gap of unknown length
* 11951 30646: contig of 18696 bp in length
* 30647 30747: gap of unknown length
* 30747 43804: contig of 13058 bp in length
* 43805 43905: gap of unknown length
* 43905 99520: contig of 55616 bp in length
* 99521 99621: gap of unknown length
* 99621 177509: contig of 77889 bp in length.
```


[illegible]

Search completed: August 2, 2002, 13:35:08
Job time: 10289 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 2, 2002, 10:39:54 ; Search time 1628.88 Seconds
(without alignments)
7780.586 Million cell updates/sec

Title: US-09-900-038a-2
Perfect score: 939
Sequence: 1 atgaattatgatcattat.....gaggagagaacaagtgc 939

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_estl:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	89.8	9.6	1101	12	CNS00EVL
c 2	87.2	9.3	1101	12	CNS0039G
c 3	77	8.2	1101	12	CNS00EVL
c 4	71.6	7.6	1248	12	B11336
5	69.6	7.4	893	12	CNS013XE
c 6	68.8	7.3	928	12	CNS0102F
7	68.4	7.3	928	12	CNS00DKY
8	68.4	7.3	1101	12	CNS0039G
c 9	68.2	7.3	1101	12	CNS017RE
10	67.2	7.2	1001	12	CNS01400
c 11	67.2	7.2	1101	12	CNS00EVL
c 12	66.8	7.1	1025	12	CNS01472
c 13	66.2	7.1	1169	12	CNS06KHQ
c 14	65.4	7.0	945	10	BM358135
15	65.4	7.0	1092	12	CNS020K7
16	64.2	6.8	910	12	CNS01G8P
17	63.8	6.8	1101	12	CNS00EPO

18	63.6	6.8	987	12	CNS014PQ
c 19	63.4	6.8	1101	12	CNS001FB
c 20	63	6.7	782	9	BB622182
c 21	62.2	6.6	867	12	CNS0758G
22	62	6.6	1101	12	CNS0021J
c 23	61.8	6.6	756	12	BH575203
24	61.6	6.6	777	12	CNS025WB
25	61.6	6.6	916	12	B09084
26	61.6	6.6	945	12	CNS04DOK
27	61.6	6.6	994	12	CNS04NOJ
28	61.4	6.5	1085	12	CNS016YR
c 29	61.2	6.5	938	12	CNS006TJ
c 30	61.2	6.5	1101	12	CNS0175J
31	60.8	6.5	1101	12	CNS008X3
32	60.8	6.5	1101	12	CNS00YWL
c 33	60.8	6.5	1101	12	CNS05AB2
34	60.8	6.5	1200	12	CNS016CO
35	60.6	6.5	1190	12	CNS02ON7
36	60.4	6.4	734	12	CNS010MP
37	60.4	6.4	849	9	AL578920
c 38	60.4	6.4	1101	12	CNS014XJ
39	60.2	6.4	886	12	BH177277
40	60.2	6.4	886	12	CNS07JUX
41	60.2	6.4	996	12	CNS00FUH
42	60.2	6.4	1101	12	CNS003BD
43	60.2	6.4	1101	12	CNS00EO7
c 44	60.2	6.4	1190	12	CNS02ON7
c 45	60	6.4	1101	12	CNS00BO1

ALIGNMENTS

RESULT 1

CNS00EVL

LOCUS

DEFINITION

CNS00EVL 1101 bp DNA linear GSS 04-JUN-1999
Drosophila melanogaster genome survey sequence T7 end of BAC:
BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AL069706.1 GI:4949849

GSS.

fruit fly.

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

Genoscope.

1 (bases 1 to 1101)

Direct Submission

Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :

BP 191 91006 EVRY cedex - FRANCE (E-mail : seqrefgenoscope.cns.fr

- Web : www.genoscope.cns.fr)

Determination of this BAC-end sequence was carried out as part of a

collaboration with the Berkeley Drosophila Genome Project (BDGP).

The BDGP is constructing a physical map of the Drosophila

melanogaster genome using these BACs. For further information

please see <http://www.fruitfly.org> The BDGP Drosophila

melanogaster BAC library was prepared by Kazutoyo Osoegawa and

Aaron Mamoser in Pieter de Jong's laboratory in the Department of

Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,

NY. The library is named RPCI-98 and was constructed by partial

ECORI digestion of Drosophila DNA provided by the BDGP from the

isogenic strain y2; cn bw sp, the same strain used for the BDGP's

PI and EST libraries. A more detailed description of the library

and how to order individual BAC clones, the entire library, or

filters for hybridization from the BACPAC Resource Center can be

found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

Location/Qualifiers

1..1101

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/clone_lib="RPCI-98"

FEATURES

source

[illegible]


```
Matches 233; Conservative 0; Mismatches 289; Indels 0; Gaps 0;
QY 340 gatttctcagcaacttaataagaattgatagaccacaaaggaatttagtatataaacaa 399
      ||||| || || || || || || || || || || || || || || || || || ||
Db 716 GATTTTAAAAAAAATTTAAAAAGATATTATATAAAAAATTTATAAAAAATNAAA 657
QY 400 cgagaaagtaataataacttaactaatactatgatacgggaagatgtttagatgatct 459
      || || || || || || || || || || || || || || || || || || || ||
Db 656 ATAAATATATAATAAATAATTAATAATAATAAAAAAATATAATAATAAAAAAATAAT 597
QY 460 atacttgcccccacacgctggcgtaaaaaaagaagtttctgataaagtttaaggatatt 519
      || || || || || || || || || || || || || || || || || || || ||
Db 596 ATATATTATATAAAAAATAAAAAATTTATATAATAATAATAATAATAATAATAATA 537
QY 520 agagatttagtcctgctgaagattatgtttgcgaataagggagctcgtgctgatttc 579
      || || || || || || || || || || || || || || || || || || || ||
Db 536 AAAAAATTTAATAAAAAATATATATTTATATTTTAAAAAATTTAAAAATATNTTT 477
QY 580 aaaaatcggttactcaataaagcttctcacagtagattaaacgagagaatggaatatca 639
      || || || || || || || || || || || || || || || || || || || ||
Db 476 NNTTTTNTAAAAAATAAAAAATTTNNNTTTTAAANTATTTTATAAATTTATAA 417
QY 640 caaaccaataaagtttaagcaatatatttactcagctattttcaagagattttataaaga 699
      || || || || || || || || || || || || || || || || || || || ||
Db 416 ATATTATAAAAAATTTAAAAAATAAAAAATTTAAATAATTAATATTTATTAATAA 357
QY 700 aaatcttatattgatatacaaaaattactaaattacttctcaagagtagctgataaaga 759
      || || || || || || || || || || || || || || || || || || || ||
Db 356 ATAAATTTAATAAAAAATAAAAAATTTAAATAATTAATTTTAAAAAATTTAATA 297
QY 760 cgcatactcagcagagctctcctaaataatttttgagctaaaaatctaccctagttact 819
      || || || || || || || || || || || || || || || || || || || ||
Db 296 AANNATANTTTAAAAAATTTAAATAATTTAAATAATTAATTTTAAAAAATAATAA 237
QY 820 attagaaaaactatatatttgtttatattttatctatttataccttaagctct 861
      || || || || || || || || || || || || || || || || || || || ||
Db 236 TTTTAAAAATTAATTTTAAATTAATTTATTTTATNTAAANNTTT 195
```

```
RESULT 5
LOCUS   CNS013XE                               893 bp   DNA       linear   GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN10102 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION AL103436
VERSION   AL103436.1 GI:5615047
KEYWORDS  GSS:
SOURCE    fruit fly.
ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 893)
AUTHORS   Genoscope.
TITLE     Submitted (23-JUL-1999) Genoscope - Centre National de Sequençage :
JOURNAL   BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT   Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source            1..893
                     /organism="Drosophila melanogaster"
                     /plasmid="pBeloBAC11"
                     /db_xref="taxon:7227"
                     /clone_lib="DrosBAC"
```

```
/clone="BACN10102"
/note="end : SP6"
BASE COUNT 372 a 21 c 33 g 308 t 159 others
ORIGIN
Query Match 7.4%; Score 69.6; DB 12; Length 893;
Best Local Similarity 36.5%; Pred. No. 0.034;
Matches 295; Conservative 84; Mismatches 426; Indels 4; Gaps 2;
QY 127 agagggtatttaagaacaattcttaacagaataattcagttgtgagataataagaataaaatc 186
      || || || || || || || || || || || || || || || || || || || ||
Db 23 AAAAAATNTTCTCTATTATHTATATWMTATATMTCTTTTMMATTAATTTAAATTTT 82
QY 187 ttgcttaataagaagaataattggttttagcatcaagtttgaacaaagcgctgaaattctc 246
      || || || || || || || || || || || || || || || || || || || ||
Db 83 TATATATATWTATTATATAMAMATATAAAMATATAAAAAAAMMMAMAMATTTATAAAAAAT 142
QY 247 aaggggagaaatatatttttagaagatgctgtagatatttcatatcccaagtagatttgat 306
      || || || || || || || || || || || || || || || || || || || ||
Db 143 TWAMTTAAATTTWATTTATATAAAATTTAMATTTATAMAAATAAAATTTMATAAAAA 202
QY 307 aagcaaaatcgtttctatggaggagaaattcattggatttctcagcaactcctaagaaatg 366
      || || || || || || || || || || || || || || || || || || || ||
Db 203 AAAAAAMAMTAAAAATAAAAAATATTTTMMATTTTMMATTTATTTTAAAAAATAN 262
QY 367 atagaccaaaaggaaatttagtatataaacaacgagaaagtaataaataatacttaact 426
      || || || || || || || || || || || || || || || || || || || ||
Db 263 CTAAAAATTTATMTAMAAAAATAAAATAAAAAATAATAAMAMAMAMTNTTMMATTTTAAAM 322
QY 427 aatgatatacgggaagatgttattgaatagatctacttgcacccacacgcgtggcgcta 486
      || || || || || || || || || || || || || || || || || || || ||
Db 323 MAAMATANTNTNAAAAATATNTTNAAMTAAAAAAMTAT--MMAAMMAAATATTATATATM 379
QY 487 aaaaagaaagtttctcgaagaattgaatggagatatagagatttagcactgttgaagattat 546
      || || || || || || || || || || || || || || || || || || || ||
Db 380 MTAATTAAMTTAAAAAATAAAAAATTAATTTMATTTMAATTAATAATTTTAAATTTT 439
QY 547 gatttgcgaataagaggagctcgtgctgatttcaaaatcgcttactcaataaagctact 606
      || || || || || || || || || || || || || || || || || || || ||
Db 440 ATWTTTTTTCATANCACCTCCTATTTCATNNAACACCACCAATTTCTCCCTATTATTTT 499
QY 607 ttacagtatagattaaacagagatggaatatcacacaaccaataagtttaagcaatatatt 666
      || || || || || || || || || || || || || || || || || || || ||
Db 500 TTWATCTATATTTTMAAAAMTATTTAANAATAAAAAATTTAAAMATTTTMTTAAATTTT 559
QY 667 tactcagctattttcaagaatttttataagaagaaatcttataatgatacaaaaaatt 726
      || || || || || || || || || || || || || || || || || || || ||
Db 560 AATTTAAATAAAAAATAATATTATAAAAAAATAAAAAATTAATAAAAMWMAAAWATTAAT 619
QY 727 actaatctatttc-aagagtatgtgataaagaacgctatatactcagcagagctctctaa 785
      || || || || || || || || || || || || || || || || || || || ||
Db 620 TGAATTTTTTTTWWAATAAAARWTCGAARGGKGGGGKTTTKGGATGRRRCAATA 679
QY 786 atatttggagctaaaatctacccttagtattactatctagaaaactatataatttggttata 845
      || || || || || || || || || || || || || || || || || || || ||
Db 680 AAAAAATAAATAAATAATTAATTTTATTTTATTTTAAAAAATTTTAAATTTTTRAT 739
QY 846 ttatactttaagctcccttggtagagggttattataataatgataataatttagt 905
      || || || || || || || || || || || || || || || || || || || ||
Db 740 KRAAARDWAATKRAATRTKRTKRDWTKDNGNTGRAAAATTTTTRDRRTTWTTRDKKRA 799
QY 906 actgaaattgttggaggagagaaacaaa 934
      || || || || || || || || || || || || || || || || || || || ||
Db 800 KTDWNCATTTTDRRAATTKRAAAAAAAR 828
```

```
RESULT 6
LOCUS   CNS0102F/c
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN03B09 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
```

```
CNS0102F
Drosophila melanogaster genome survey sequence T7 end of BAC
BACN03B09 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
```

```
ACCESSION AL098433.1 GI:5610044
VERSION AL098433.1
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 928)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelobAC11.
FEATURES             Location/Qualifiers
     source           1..928
                     /organism="Drosophila melanogaster"
                     /plasmid="pBelobAC11"
                     /db_xref="taxon:7227"
                     /clone_lib="DrosBAC"
                     /clone="BACN03B09"
                     /note="end : T7"
BASE COUNT  169 a   128 c   100 g   241 t   290 others
ORIGIN

Query Match      7.3%; Score 68.8; DB 12; Length 928;
Best Local Similarity 24.5%; Pred. No. 0.046;
Matches 126; Conservative 156; Mismatches 232; Indels 0; Gaps 0;

Qy 6 ttaagtagtattatgctcggtataataatgagcctttaaattatgtgagagattcagtaga 65
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 924 TAAGAKGWAAGWRRWAATDRWATAWTAKWRGAGTRWATARKATKAGKAGAKRWAA 865

Qy 66 atctattataaacaacgcttactgatttttgagttcataatttcattgattccaaag 125
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 864 RWATAKAKWGTWCAKWTADKDKTGGTGTGWTGTRGTRGKRRRAGGDTGRGDRTR 805

Qy 126 tagaggtgatttaagaacaattcttaacagaataattcagttgtagatataagaataaaaaat 185
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 804 TRRAGKAGRWAKARAKTAWAADNGARWAWAAAWADATGAGATAMWATAWTATAAKATA 745

Qy 186 ctgcttaataagaagaaatattggttttagcatcaagtttgacaagcggtgaaatttc 245
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 744 TTAWAWRRDRARATGTGKWRGTGTTAKAGATKGGWDDTRADAKKAAGGTTATAKTATA 685

Qy 246 taegggagaatattttttagaattggtgctgtagtatatttcataatccaaagtagattga 305
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 684 TGTAKDRARAKWKAATAKAKAGATATRAAWATGADAWGATWTGKKTAGAKTTGTAGW 625

Qy 306 taagcaaatctgtttttagaggaaaattcattggtattcttcacgaactctaatagaatt 365
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 624 TGRKWTDKRRKTKDRATGWRRAATAWAGAADDAWDRGWRAGTGCAGRRRRDGRWRARA 565

Qy 366 gatagacccaaaagaaatttagtataaacaacgagaaagtaataaaaataatactaaac 425
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 564 DTRKAGADRRAAAAAAGTGTGADAADKRAARGKKGDKGKRRKTKGKRAATGGWRGAG 505

Qy 426 taatgatatacgggaagtgttattgaatagatctctattctgcccccacccaacgtggtcgt 485
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 504 RAKDGRKRGWKKRKAADKTKTWDTRTDKAKAKATGAKDGAKGDKGAKRWAGDKRKKGGG 445

Qy 486 aaaaaagaagttttcagataagttcaatgggatat 519
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 444 AGATATRTGWDWDAWTTGAWTTWATGRTATAT 411
```


TITLE An integrated analysis of the genetics, development, and evolution of the cotton fiber
JOURNAL Unpublished (2000)
COMMENT Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Total High Quality bases = 126
Seq primer: TAAATGAGCTCACTATAGG
High quality sequence start: 3
High quality sequence stop: 842.
Location/Qualifiers
1. .945
/organism="Gossypium arboreum"
/strain="AXA"
/cultivar="8400"
/db_xref="taxon:29729"
/clone="GA_Ea0005115r"
/clone_lib="Gossypium arboreum 7-10 dpa fiber library"
/tissue_type="Fibers isolated from bolls harvested 7-10 dpa"
/lab_host="E. coli"
/note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"

FEATURES
source

BASE COUNT 484 a 57 c 18 g 368 t 18 others
ORIGIN

Query Match 7.0%; Score 65.4; DB 10; Length 945;
Best Local Similarity 46.1%; Pred. No. 0.16;
Matches 343; Conservative 0; Mismatches 389; Indels 12; Gaps 4;

QY 159 ttacgttgtagataagataaaaaattcgtttaaagaagaaatatatggtttgagatc 218
DB 204 TTCACTTTTCTTTTCTTTTGGAAATTTTATTTTACTTCATATATCATTTTAACTTT 263
QY 219 aggtttgaacagcggtgaaattcgaaggagagataatattttgagatgagctga 278
DB 264 TTTTCTTTACTA----TCAGATTTTNCAAAAAATACATTTATTTTATATTTT 317
QY 279 tgatatttcataatcgaagtagatttgataagcaaatctgtttatgagagaaatcatt 338
DB 318 TTTTTCACCAATAATTTATTAATAATAAACTATTTTAAATATTTTATTTATTTT 377
QY 339 ggaattctcagcaactctagaattgatagaccacaaagaaatttagatataaaca 398
DB 378 TTNTTTNNNAATAAATGAAATTTATTAATAATAATAATAATAATAATAATAA 437
QY 399 acgagaagaataaataatacttaactaatgatatacgcgaagattgttattgaatagatc 458
DB 438 ACAAAAAATTAATAACAT--TAAATAATAAAAAAACAATAATAATAATAATAA 495
QY 459 tatactgcccacccacacggtggtgtaaaaaagaaagtgttcgataaagttaatgggata 518
DB 496 AAAAAAATAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 555
QY 519 tagagatttagctacgttgaagattgatatttgcaataagagagctcgctgattt 578
DB 556 AAAAAAATAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 615
QY 579 caaaatcggtctactcaataaagttacttttacagtagatagataaacgagaatggaatc 638
DB 616 -AANTTATATAATAAATAAATCATATAAAAAAATAATAATAATAATAATAATAA 674
QY 639 acaaaccaataagtttaagcaatataatttactcagctattttacaagatttttataaaga 698
DB 675 ATAAAAATAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 734
QY 699 aaaaattctatgatcacaaaaataacttaacttctcagagatttgatgaagaaga 758
DB 735 TTNTNTNAAANTTAATTTATATATATTAATAATAATAATAATAATAATAATAA 794

QY 759 acgtatactcagcaagagctctctaaatttttgagctaaatctacccttagtattac 818
DB 795 A---TAAATAA 851
QY 819 tattagaaaactatataatttggtttatattatacttaagctcccttggtgaggggtt 878
DB 852 AATATTATAAATTTNTTTTATTAATAAATTTNTNATAATATTNATAATATATAAAT 911
QY 879 attaataaatgatataataatttt 902
DB 912 ATAATTTAAATATATATATATTT 935

RESULT 15

CNS020K7 1092 bp DNA linear GSS 12-MAY-2000
Tetraodon nigroviridis genome survey sequence T7 end of clone
222L11 of library G from Tetraodon nigroviridis, genomic survey
sequence.

ACCESSION AL175696.1 GI:7813753
VERSION
KEYWORDS GSS: genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis.

REFERENCE 1 (bases 1 to 1092)
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.
TITLE Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1092)
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 1092)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
COMMENT This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.

FEATURES
source

1..1092
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="222L11"
/clone_lib="G"
/note="Genoscope sequence ID : COAG222CF06LP1-end : T7"
BASE COUNT 383 a 169 c 165 g 262 t 113 others
ORIGIN

Query Match 7.0%; Score 65.4; DB 12; Length 1092;
Best Local Similarity 36.5%; Pred. No. 0.16;
Matches 153; Conservative 59; Mismatches 207; Indels 0; Gaps 0;

QY 143 aattcttaacagaatattcagttgtagataataagaataaaattctgtcttaagagaaa 202
DB 626 AWWWWAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 685
QY 203 atattggttttagcatcaaatgttgacaacagcggtgaaattttctaaagggagatatatt 262
DB 686 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 745


```
Qy 263 ttgaatggatgctgatgatatttcacatccaaagtagatttgataagcaaatcggttta 322
Db 746 AWWKAWTTTWTTTTWTAWWWWWWWAAAAAANWKKWWTTTWTATTTTAAWWATWAWWTTWTA 805

Qy 323 tggaggaaaatttcattggattctcagcaactctctaataagattgataagccaaaaaggaa 382
Db 806 TTTTAAAWAAAAANWTTTAWTTAAWAAAAAANWAAATAATTTTAAAWAWTTTAAWAAA 865

Qy 383 atttagtatataaacaacgagaagtaataaaatatataacttaactaatgatatacggaaga 442
Db 866 TTWAAATWTAATAAAAAAATAAAAAAANWAAWTTTWTTTTWTWAWWTTTAAATWTTT 925

Qy 443 tgttattgaatagatctatacttgccaccaccaacgctgctgctaaagaaagatttcg 502
Db 926 TTTTWTTTTWTWTTTWTAAWAAAAAANWAAATAATAAANWAAWTTTAAATTTATTT 985

Qy 503 ataagttaatggagatatagagatttagtacctgttgagattatgatatttgcaataaga 561
Db 986 TTAATAAAAAAANWAAATAAATAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1044
```

Search completed: August 2, 2002, 12:56:53
Job time: 8219 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 2, 2002, 12:22:04 ; Search time 210.05 Seconds
(without alignments)
7675.233 Million cell updates/sec

Title: US-09-900-038A-2

Perfect score: 939

Sequence: 1 atgaattatagatcatatt.....gaggagagaacaaagtgc 939

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_032802.*

1: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	70.2	7.5	15548	24	ABL34155
C 2	68.6	7.3	5371	24	ABL34294
C 3	68.2	7.3	14024	20	AAK06749
C 4	68.2	7.3	14024	21	AAZ56386
C 5	67	7.1	4915	20	AAK13062
C 6	64.8	6.9	906	21	AAK13062
C 7	64.8	6.9	11474	21	AAK53723
C 8	64	6.8	7597	24	AAK53720
C 9	63.8	6.8	1170	21	AAK53728
C 10	63.8	6.8	11474	21	AAK53720
C 11	63.4	6.8	5154	20	AAK13138
C 12	63	6.7	5689	22	AAK5384
C 13	63	6.7	5689	22	AAK5384
C 14	60.4	6.4	912	21	AAK53727
C 15	60.4	6.4	7814	22	AAK53727
C 16	60.2	6.4	12237	24	ABL34358
C 17	59.8	6.4	700	22	AAH93026
C 18	59.6	6.3	6286	22	AAK46591
C 19	59	6.3	8370	22	AAK46713
C 20	58.8	6.3	2118	23	AAK51313
C 21	58.8	6.3	2148	23	AAK52777
C 22	58.8	6.3	4054	20	AAK13188
C 23	58.6	6.2	6474	19	AAV52288
C 24	58	6.2	6109	24	ABL32326
C 25	58	6.2	6109	24	AAK561077
C 26	57.8	6.2	3991	22	AAK16633
C 27	57.8	6.2	6992	21	AAK60930
C 28	57.8	6.2	9539	22	AAK45346
C 29	57	6.1	18183	22	AAK46280
C 30	56.6	6.0	15950	21	AAK45260
C 31	56.6	6.0	15950	21	AAK30357
C 32	56.6	6.0	19124	18	AAK72882
C 33	56.6	6.0	19124	21	AAK98287
C 34	56.4	6.0	1671	13	AAQ24134
C 35	56.4	6.0	37973	24	ABL34197
C 36	56	6.0	18218	24	ABL33949
C 37	55.6	5.9	5883	24	ABL33733
C 38	55.6	5.9	19124	18	AAK72882
C 39	55.6	5.9	19124	21	AAK98287
C 40	55.4	5.9	6154	24	AAK561112
C 41	55.2	5.9	700	22	AAK93026
C 42	55	5.9	20420	22	AAK73165
C 43	54.8	5.8	5930	24	ABL32517
C 44	54.8	5.8	6419	24	ABL32267
C 45	54.8	5.8	11155	24	ABL32605

ALIGNMENTS

RESULT 1
ABL34155/c
ID ABL34155 standard; DNA; 15548 BP.
XX
AC ABL34155;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 2128.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianemic; cytostatic; neuroproliferative;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antineuritic; antitumor; antidiabetic; antipsoriatic;
KW antineuritic; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP07537.
XX
PR 30-JUN-2000; 2000DE-1032529.
XX
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX

LPS core biosynthe
Enterococcus faeca
Chemically pretrea
Tumour suppressor
Campylobacter jeju
Tumour suppressor
Human immune syste
Human inflammatory
Tumour suppressor
Tumour suppressor
Enterococcus faeca
Enterococcus faeca
Streptococcus pneu
Human immune syste
Human gene regulat
Human novel protei
Nucleotide sequenc
Chemically pretrea
Tumour suppressor
DNA encoding enzym
Nucleotide sequenc
Plasmodium var-7 g
50 kD subunit of S
Human immune syste
Human immune syste
Human immune syste
Plasmodium var-7 g
Human gene regulat
Human inflammatory
Human immune/haema
Human immune syste
Human immune syste
Human immune syste

Matches 198; Conservative 0; Mismatches 173; Indels 6; Gaps 2;

Qy 13 atcattatgtcggtataataagagccctttaaatttatgtgagagattcagtagaatctata 72
|| ||||| || ||||| || || || ||||| ||||| ||||| ||

Dd 100 ataattatgccggtttacaatgggc---caaaactataatctcatcagtagaatcaatt 156

Qy 73 ttaaatcaaacgcttactagtttttgagttccaattgttcattgataatccaagtagagg 132
|| ||||| | ||||| | || ||||| || ||||| || || || |

Dd 157 atacatcaatttatcaagatttg---tttgtatatcattgacgattgtagaccgat 213

Qy 133 gatttaaagcaattctttaacgaataatttcagttgtagataatagaataaaaaatttgctt 192
|| | || | || | || | || | || | || | || | || | || | || |

Dd 214 gatcacatttcttaatacaacagtcgatacaaaaacaatcagaaataagaattatgcgt 273

Qy 193 aatgaagaaaataattggtttagcatcaagtttgaccaagcgggtgaaaatttctaaggga 252
|| | || | || | || | || | || | || | || | || | || | || |

Dd 274 aacaagcaaatctaggtgttcgaaagtcgaattatggaatagaataatggcccagggg 333

Qy 253 gaatatatttttgaatggatgctgatgatatttcattatcccaagtagattgataagcaa 312
|| ||||| || ||||| ||||| ||||| || || || || || || ||

Dd 334 aaatatattttttgtgatgcggtgatttggcagcagagaaaaatttagcgcgtcaa 393

Qy 313 attcgttttatggaggaaaatttcattgatttttcagcaaccttaataagaattgatagac 372
|| | || | || | || | || | || | || | || | || | || | || |

Dd 394 atcgaagtgttaataatgaatgtgatgtggtgttcttaattattatgttatagat 453

Qy 373 caaaaaggaaaatttagt 389

Dd 454 aacaatagaataattgt 470

RESULT 4

AZ56386 ID AZ56386 standard; DNA; 14024 BP.

XX XX AZ56386;

XX XX 17-MAR-2000 (first entry)

XX XX Escherichia coli flagellin protein nucleotide sequence SEQ ID NO:56.

DE DE Flagellin; fliC; antigen; detection; ds.

KW KW Escherichia coli.

XX OS WO961458-A1.

XX PN 02-DEC-1999.

PD PD 21-MAY-1999; 99WO-AU00385.

XX PF 21-MAY-1998; 98AU-0003634.

PR PR (UNSY) UNIV SYDNEY.

XX PA Reeves PR, Wang L;

XX PI WPI; 2000-072598/06.

XX DR Novel nucleic acid molecule useful for the detection of flagellated bacterial strains in food, faeces, etc. .

XX PS Claim 3; Page 229-233; 245pp; English.

XX CC AZ56331 to AZ56398 represent nucleic acid molecules (I) encoding all or part of an Escherichia coli flagellin protein except a protein expressed by E. coli H1, H7, H12 or H48 type strains. The present invention also describes a method of detecting the presence of E. coli of a particular H serotype in a sample, comprising specifically hybridising a nucleic acid, preferably at least a pair, derived from a flagellating gene, specific for a particular flagellin gene associated with the H serotype, to any E.coli in the sample which contain the gene, CC and detecting any hybridised molecules, identifying the presence of that

CC	serotype in the sample. (I) are useful for: (1) detecting the presence	
CC	of E. coli of H serotype in a sample by hybridising at least one or a	
CC	pair of (I) to any E. coli in the sample and detecting the hybridised	
CC	nucleic acid molecules; and (2) for detecting the presence of both O	
CC	and H-serotypes of E. coli by hybridising at least one or a pair of (I)	
CC	to any E. coli present in the sample and detecting the hybridised	
CC	nucleic acid molecules. (I) is particularly useful for detecting the	
CC	combination of O and H antigen. Hybridised (I) when using at least one	
CC	(I) is detected by southern blot analysis and, when using a pair of (I),	
CC	is detected by polymerase chain reaction (PCR). AAZ56399 to AAZ56420	
CC	represent primers used in the exemplification of the present invention.	
XX		
XX	Sequence 14024 BP; 4546 A; 2263 C; 2832 G; 4383 T; 0 other;	
SO		
	Query Match 7.3%; Score 68.2; DB 21; Length 14024;	
	Best Local Similarity 52.5%; Pred. No. 0.00084;	
	Matches 198; Conservative 0; Mismatches 173; Indels 6; Gaps 2;	
Qy	13 atcattatgtcgggtatataaagagcctttaaatatgtgagagattcagtagaattcata 72	
Db	100 ataattatgccggtttacaattggggc---caaaactataatctcatcagtagaatcaatt 156	
Qy	73 ttaaatcaaacgcttaactgattttgagttcotaattgtcattgataatccaagttagaggt 132	
Db	157 atacatcaattctatcaagattttg---ttttgtatcatcattgcagattgtgaccgat 213	
Qy	133 gatttaagaacaattcttaacagaatatttcagttgtagataatagaataaaaaatttgctt 192	
Db	214 gatacattttcttaatacaacagtcgatacaaaaacatcagaaaataagaatttcgt 273	
Qy	193 aatgaagaaaaatatggtttagtcataagtttgacaaaagcgttgaaaattctctaaaggga 252	
Db	274 acaagcaaaatttaggtgttcgaaagtcgaattatggaaatagaatggccacagggg 333	
Qy	253 gaatatatttttagaattggtcgtgatgatatttcataatccaagtagattgtgataagcaa 312	
Db	334 aaatatattttttgtgtagcggatgatttggcagcagagaaaaattagagcggtcaa 393	
Qy	313 attcgttttatggaggaaaaattcattggattttcagcaactctaatagaattgatagac 372	
Db	394 atcgaagtgttaataatgaatgtgtagatggtgttatgttcttaattattgtttatagat 453	
Qy	373 caaaaaggaaaatttcagt 389	
Db	454 acaatatgaataattgt 470	
RESULT	5	
AXX13062/c		
ID	AXX13062 standard; DNA; 4915 BP.	
XX		
AC	AXX13062;	
XX		
DT	19-MAR-1999 (first entry)	
XX		
DE	Enterococcus faecalis genome contig SEQ ID NO:125.	
XX		
KW	Enterococcus faecalis; contig; detection; Enterococcal infection;	
KW	vaccine; attenuation; computer readable medium; ds.	
XX		
OS	Enterococcus faecalis.	
XX		
PN	W098050555-A2.	
XX		
PD	12-NOV-1998.	
XX		
PF	04-MAY-1998; 98WO-US08985.	
XX		
PR	14-NOV-1997; 97US-0066009.	
PR	06-MAY-1997; 97US-0044031.	
PR	16-MAY-1997; 97US-0046655.	
XX		


```
QY 181 aaaaatttgcttaataagaagaaaataattgtgttagcatcaagttagaacaacggtgaaa 240
||||| | ||||| |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Db 178 aaaaataatccacaataagaaaacttagtgccttttaagagcaagatatgaaggtgtgaaa 237
||||| | ||||| |||| | |||| | |||| | |||| | |||| | |||| | |||| |
QY 241 atttctaaggaggagaatatattttttagaatggatgctgtgatgattatccatccaagttaga 300
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 238 gtacgaactcctctatatataatgttttttagatcctgtgatgattatttggaaactaaatgct 297
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 301 ttgtataagcaaatctgctttttagaggagaaa 332
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 298 tgtgaagagtgtataaaaaatttttagatgaaca 329
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 7
ID AAA53720/c
XX AAA53720 standard; DNA; 11474 BP.
XX AAA53720;
XX 22-DEC-2000 (first entry)
XX LPS core biosynthesis locus.
XX Biosynthetic locus; biosynthesis; lipid A biosynthesis;
KW acetyltransferase; glycosyltransferase; Beta-1,4-GalNAc transferase;
KW Beta-1,3-galactosyltransferase; alpha-2,3-sialyltransferase;
KW sialic acid synthase; CMP-sialic acid synthetase; mimic; antibody;
KW immunity; immunogen; ganglioside; ds.
XX Campylobacter jejuni OH4384.
XX Key Location/Qualifiers
FH 350..1237
CDS /*tag= a
/*product= acyltransferase
/*note= "Open reading frame 2b"
1234..2490
CDS /*tag= b
/*product= glycosyl transferase
/*note= "Open reading frame 3a"
2786..3955
CDS /*tag= c
/*product= Glycosyl transferase
/*note= "Open reading frame 4b"
4025..5068
CDS /*tag= d
/*product= Beta 1,4-GalNAc transferase
/*note= "Open reading frame 5a"
5054..5959
CDS /*tag= e
/*product= Beta 1,2- galactosyltransferase
/*note= "Open reading frame 6a"
6048..6923
CDS /*tag= f
/*product= CstII sialyltransferase
/*note= "Open reading frame 7a"
6924..7964
CDS /*tag= g
/*product= Sialic acid synthase
/*note= "Open reading frame 8a"
8020..9079
CDS /*tag= h
/*product= Sialic acid biosynthetic enzyme
/*note= "Open reading frame 9a"
9076..9741
CDS /*tag= i
/*product= CMP-sialic acid synthetase
/*note= "Open reading frame 10a"
complement (10554..11366)
CDS /*tag= j
/*product= Lipooligosaccharide biosynthetic enzyme
/*note= "Open reading frame 12a"
```

```
PN WO200046379-A1.
XX 10-AUG-2000.
XX 01-FEB-2000; 2000WO-CA00086.
XX 01-FEB-1999; 990US-0118213.
XX 31-JAN-2000; 2000US-0495406.
XX (CANA ) NAT RES COUNCIL CANADA.
XX Gilbert M, Wakarchuk WW;
PI WPI; 2000-524418/47.
DR P-PSDB; AAY97200, AAY97201, AAY97202, AAY97203, AAY97204, AAY97205,
DR AAY97206.
XX Novel glycosyltransferase polypeptides and polynucleotides useful for
PT biosynthesis of ganglioside and ganglioside mimics, as diagnostic
PT reagents and as immunogen for producing antibodies
XX
PS Claim 1; Page 86-90; 120pp; English.
XX A reaction mixture for the synthesis of a sialylated oligosaccharide
CC is useful for synthesizing sialylated oligosaccharide such as
CC ganglioside, lysoganglioside or their mimics. Glycosyltransferases
CC are useful for chemo-enzymatic synthesis of oligosaccharides,
CC including gangliosides and other oligosaccharides that have
CC biological activity. The enzymes and nucleic acids that encode them
CC are useful for studies of the pathogenesis mechanisms of organisms
CC that synthesize ganglioside mimics, such as C. jejuni and the
CC nucleic acids are used as probes to study expression of genes
CC involved in ganglioside mimetic synthesis. Antibodies raised against
CC the glycosyltransferases are also useful for analyzing the
CC expression patterns of these genes involved in pathogenesis. The
CC nucleic acids are also useful for designing antisense
CC oligonucleotides for inhibiting expression of the Campylobacter
CC enzymes that are involved in the biosynthesis of ganglioside mimics
CC that can mask the pathogens from the host's immune system. The
CC oligosaccharides are useful as diagnosing reagents or as therapeutics
CC and as immunogens for producing antibodies. Bacterial
CC glycosyltransferase can be used to catalyse the formation of
CC oligosaccharides that are identical to the corresponding mammalian
CC structures and are easier and less expensive to produce in large
CC quantity, compared to the mammalian glycosyltransferase. The
CC bacterial origin of the enzymes facilitates expression of large
CC quantities of the enzymes using relatively inexpensive prokaryotic
CC expression systems.
XX
SQ Sequence 11474 BP; 4506 A; 1495 C; 1613 G; 3860 T; 0 other;
```

```
Query Match 6.9%; Score 64.8; DB 21; Length 11474;
Best Local Similarity 51.8%; Pred. No. 0.0034;
Matches 172; Conservative 0; Mismatches 157; Indels 3; Gaps 1;
```

```
QY 1 atgaattatagatcattatgtcggtatataatgagccttttaaatattgtgagagattca 60
||||| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 5959 ATGTTTAAATTTCAATCATCTTACCAACTTATATGTGGACAATATATAGCAAGGCA 5900
QY 61 gtagaatctatattaatacaacgcgttactgtattgtgatttcataattgtcattgataat 120
||||| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 5899 ATAGAAAAGCTGTATCAATCAGACTTTTAAAGATATAGA---AATATTTAGTTGATGAT 5843
QY 121 ccaagttagagggtatttaagaacaaattcttaacagaataattcagttgtagataagaata 180
||||| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 5842 TGTGGAATCATATAGTATATAATATAGCCAAAGAATACTCTTAAMAAACAAAGAAATA 5783
QY 181 aaaaatttgcttaataagaagaaaataattgtgttagcatcaagttagaacaacggtgaaa 240
||||| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 5782 AAAATAATCCCAATCGAAAAAACTTAGTCTCTTTTAAAGAGCAAGATATGAAGGTGTGAAA 5723
QY 241 atttctaaggaggagaatatatttttagaatggatgctgtgatatttcatcatccaagttaga 300
```


XX (CANA) NAT RES COUNCIL CANADA.
 XX
 PI Gilbert M, Wakarchuk WW;
 XX
 DR WPI; 2000-524418/47.
 DR P-PSDB; AAY97202.
 XX
 PT Novel glycosyltransferase polypeptides and polynucleotides useful for
 PT biosynthesis of ganglioside and ganglioside mimics, as diagnostic
 PT reagents and as immunogen for producing antibodies
 XX
 PS Disclosure; Page 99; 120pp; English.
 XX
 CC A reaction mixture for the synthesis of a sialylated oligosaccharide
 CC is useful for synthesising sialylated oligosaccharide such as
 CC ganglioside, lysoganglioside or their mimics. Glycosyltransferases
 CC are useful for chemo-enzymatic synthesis of oligosaccharides,
 CC including gangliosides and other oligosaccharides that have
 CC biological activity. The enzymes and nucleic acids that encode them
 CC are useful for studies of the pathogenesis mechanisms of organisms
 CC that synthesize ganglioside mimics, such as C. jejuni and the
 CC nucleic acids are used as probes to study expression of genes
 CC involved in ganglioside mimetic synthesis. Antibodies raised against
 CC the glycosyltransferases are also useful for analyzing the
 CC expression patterns of these genes involved in pathogenesis. The
 CC nucleic acids are also useful for designing antisense
 CC oligonucleotides for inhibiting expression of the Campylobacter
 CC enzymes that are involved in the biosynthesis of ganglioside mimics
 CC that can mask the pathogens from the host's immune system. The
 CC oligosaccharides are useful as diagnosing reagents or as therapeutics
 CC and as immunogens for producing antibodies. Bacterial
 CC glycosyltransferase can be used to catalyze the formation of
 CC oligosaccharides that are identical to the corresponding mammalian
 CC structures and are easier and less expensive to produce in large
 CC quantity, compared to the mammalian glycosyltransferase. The
 CC bacterial origin of the enzymes facilitates expression of large
 CC quantities of the enzymes using relatively inexpensive prokaryotic
 CC expression systems.
 XX
 SQ Sequence 1170 BP; 504 A; 120 C; 171 G; 375 T; 0 other;

Query Match 6.8%; Score 63.8; DB 21; Length 1170;
 Best Local Similarity 56.8%; Pred. NO. 0.0049;
 Matches 138; Conservative 0; Mismatches 102; Indels 3; Gaps 1;

QY 42 aaattatgtgagagattccagtagaattctattataatcaaacgcttactgatttgagtt 101
 DB 42 aaattatgtgagagattgttagatagcgtttatcaatcaaaccttactaacttagaaat 101
 QY 102 cataattgtcattgataatccagtagaggtgtattaaagcaattcttaacagaatattc 161
 DB 102 cataattgtcattgattggttagcacagatgaacactcaactatattgcaaaagaatatac 161
 QY 162 agttgtagataatagaataaaatcttgccttaataagaataattggtttagcatcaag 221
 DB 162 cttaaaagataaagaataactcttt---tgataaagaataatggggtttaagttcagc 218
 QY 222 ttgtgaacaagcgggtgaataattcttaggggagaataatatttttagaattggtcgtgata 281
 DB 219 tagaataataggtagataacttttagcggggaataataataaaacaaactcaaca 278
 QY 282 tat 284
 DB 279 tat 281

RESULT 10
 ID AAA53720
 ID AAA53720 standard; DNA; 11474 BP.
 XX
 AC AAA53720;

XX 22-DEC-2000 (first entry)
 DT
 XX LPS core biosynthesis locus.
 DE
 XX Biosynthetic locus; biosynthesis; lipid A biosynthesis;
 KW acetyltransferase; glycosyltransferase; Beta-1,4-GalNAC transferase;
 KW Beta-1,3-galactosyltransferase; alpha-2,3-sialyltransferase;
 KW sialic acid synthase; CMP-sialic acid synthetase; mimic; antibody;
 KW immunity; immunogen; ganglioside; ds.
 XX
 OS Campylobacter jejuni OH4384.
 XX
 FH Key Location/Qualifiers
 FT CDS 350..1237
 FT /tag= a
 FT /product= acyltransferase
 FT /note= "Open reading frame 2b"
 FT CDS 1234..2490
 FT /tag= b
 FT /product= glycosyl transferase
 FT /note= "Open reading frame 3a"
 FT CDS 2786..3955
 FT /tag= c
 FT /product= Glycosyl transferase
 FT /note= "Open reading frame 4b"
 FT CDS 4025..5068
 FT /tag= d
 FT /product= Beta 1,4-GalNAC transferase
 FT /note= "Open reading frame 5a"
 FT CDS 5054..5959
 FT /tag= e
 FT /product= Beta 1,2- galactosyltransferase
 FT /note= "Open reading frame 6a"
 FT CDS 6048..6923
 FT /tag= f
 FT /product= CstII sialyltransferase
 FT /note= "Open reading frame 7a"
 FT CDS 6924..7964
 FT /tag= g
 FT /product= Sialic acid synthase
 FT /note= "Open reading frame 8a"
 FT CDS 8020..9079
 FT /tag= h
 FT /product= Sialic acid biosynthetic enzyme
 FT /note= "Open reading frame 9a"
 FT CDS 9076..9741
 FT /tag= i
 FT /product= CMP-sialic acid synthetase
 FT /note= "Open reading frame 10a"
 FT CDS complement (10554..11366)
 FT /tag= j
 FT /product= Lipooligosaccharide biosynthetic enzyme
 FT /note= "Open reading frame 12a"
 XX
 PN WO200046379-A1.
 XX
 PD 10-AUG-2000.
 XX
 PF 01-FEB-2000; 2000WO-CA00086.
 XX
 PR 01-FEB-1999; 99US-0118213.
 PR 31-JAN-2000; 2000US-0495406.
 XX
 XX (CANA) NAT RES COUNCIL CANADA.
 XX
 PI Gilbert M, Wakarchuk WW;
 XX
 DR WPI; 2000-524418/47.
 DR P-PSDB; AAY97200, AAY97201, AAY97202, AAY97203, AAY97204, AAY97205,
 DR AAY97206.
 XX
 PT Novel glycosyltransferase polypeptides and polynucleotides useful for

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 2, 2002, 10:47:14 ; Search time 49.3 Seconds
(without alignments)
4678.496 Million cell updates/sec

Title: US-09-900-038A-2

Perfect score: 939

Sequence: 1 atgaattatagatcattat.....gaggagagaacaagaagtgc 939

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA:*

1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*

2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*

3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*

4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*

5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*

6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	56.6	6.0	19124	2	US-08-487-826B-13
C 2	55.6	5.9	19124	2	US-08-487-826B-13
C 3	51	5.4	854	4	US-08-998-416-534
C 4	50	5.3	860	4	US-08-998-416-287
C 5	50	5.3	6124	4	US-08-213-419B-3
C 6	49.4	5.3	7218	1	US-08-232-463-14
C 7	48.8	5.2	1882	1	US-08-257-073-12
C 8	48.8	5.2	1884	1	US-08-257-073-8
C 9	48.6	5.2	615	4	US-08-998-416-186
C 10	47.6	5.1	844	3	US-08-961-083-195
C 11	47.4	5.0	837	4	US-08-998-416-288
C 12	46	4.9	3701	4	US-08-845-258-10
C 13	46	4.9	3701	4	US-08-990-571-10
C 14	46	4.9	3701	4	US-08-723-142A-10
C 15	45.6	4.9	636	4	US-08-998-416-1137
C 16	45.6	4.9	665	2	US-08-883-795A-36
C 17	45.6	4.9	6152	4	US-08-973-462-1
C 18	45.6	4.9	8920	2	US-08-446-855A-1
C 19	45.6	4.9	8920	4	US-09-150-741-1
C 20	45.4	4.8	658	4	US-08-998-416-595
C 21	44.4	4.7	837	4	US-08-998-416-288
C 22	44.2	4.7	732	4	US-08-998-416-1036
C 23	44	4.7	724	4	US-08-998-416-683
C 24	43.8	4.7	688	4	US-08-998-416-972
C 25	43.8	4.7	1684	4	US-08-936-165A-259
C 26	43.4	4.6	782	4	US-08-998-416-224
C 27	43	4.6	767	4	US-08-998-416-472

C 28	42.8	4.6	663	4	US-08-998-416-191	Sequence 191, App
C 29	42.4	4.5	663	4	US-08-998-416-191	Sequence 191, App
C 30	42.2	4.5	1164	3	US-08-755-587-32	Sequence 32, Appl
C 31	42.2	4.5	1850	3	US-08-617-860B-32	Sequence 32, Appl
C 32	42.2	4.5	4098	2	US-08-605-106-4	Sequence 4, Appl
C 33	41.8	4.5	8654	1	US-08-920-812-6	Sequence 6, Appl
C 34	41.8	4.5	8654	1	US-08-920-827-6	Sequence 6, Appl
C 35	41.8	4.5	8654	1	US-08-921-177-6	Sequence 6, Appl
C 36	41.8	4.5	8654	1	US-08-362-577C-6	Sequence 6, Appl
C 37	41.8	4.5	8654	2	US-08-920-828-6	Sequence 6, Appl
C 38	41.6	4.4	677	4	US-08-998-416-1092	Sequence 1092, Ap
C 39	41.6	4.4	782	4	US-08-998-416-224	Sequence 224, App
C 40	41.6	4.4	821	4	US-08-998-416-541	Sequence 541, App
C 41	41.6	4.4	854	4	US-08-998-416-534	Sequence 534, App
C 42	41.6	4.4	860	4	US-08-998-416-287	Sequence 287, App
C 43	41.6	4.4	872	4	US-08-998-416-487	Sequence 487, App
C 44	41.6	4.4	3926	2	US-08-731-722-1	Sequence 1, Appl
C 45	41.6	4.4	3926	2	US-08-731-722-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-487-826B-13/c
; Sequence 13, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellem, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israel, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19124 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: Single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-487-826B-13

Query Match

6.0%; Score 56.6; DB 2; Length 19124;


```
Db 15857 AAAAAATAAATAAAAAAATTAATTAATAAAAAAATAAAAAAATTAATAAGAA 15916
QY 634 ataccacaaacaaagtttaagcaaatatatttactcagctattttacagatttttat 693
Db 15917 ATAAAAAATAAATAAAAAATTAATTAATAAAAAAATAAAAAAATTAATTAACATGC 15976
QY 694 aagaaataat-cttatattgatatacaaaaaattactaatttactttccaagagatgtgat 752
Db 15977 ACATATACATACATATATATATATATATATATATATATATATATATATATATATATAT 16036
QY 753 aagaaacgctatactactcagcaagcgtctctaaataattttgagctaaaaactaccocctag 812
Db 16037 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 16096
QY 813 tattactattagaaactatatatttggttattatttatttacttttaagctccctgggttag 872
Db 16097 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 16156
QY 873 gaggttatttaataaagatgatatatttattttagtactgaaattgttt 918
Db 16157 TACCTAAATACATATATATATATATATATATATATATATATATATATATATATATAT 16202
```

```
RESULT 3
US-08-998-416-534/c
; Sequence 534, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippsen, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jurgen
; APPLICANT: Knechtle, Philipp
; APPLICANT: Rebschung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPHII
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6239264artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: No. 6239264th Carolina
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 24-DEC-1997
; APPLICATION NUMBER: US/08/998,416
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 534:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 854 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
```

```
; ORGANISM: PAG1372UP
US-08-998-416-534
Query Match 5.4%; Score 51; DB 4; Length 854;
Best Local Similarity 45.1%; Pred. No. 0.0072;
Matches 265; Conservative 0; Mismatches 320; Indels 2; Gaps 2;
QY 41 taaattatgtgagagattccagtagaatcttatataaatacaacgcttactgattttgagt 100
Db 649 TAAATAAATAGAAACCATAGTTAATGATCAATAAAGAAATGGAATTAATTTGTGGCA 590
QY 101 tcaaatgttcattgataatcccaagtagagtgattttaaagcaattcttcaacagaatatt 160
Db 589 TCTTAATTTTATTTAATTTAATTTGATTATATCTATTAACATATAAACATATTTTAAATGCTT 530
QY 161 cagttgtagataatagataaaaaactcttgcttaagagaaaaatatttggttttagcatacaa 220
Db 529 ATAAATAAATAAGAAATTTACTTTAGAAATATTTAATAATAGTATTTAATTTAATTTTAA 470
QY 221 gtttgaacaaagcgtgaaaaatttctaaggagagaatatatttttagaatgagatgctgatg 280
Db 469 TATTAAATATACCATTTTATTAAATAAATAGATTATTAAAGTTTATTAAATATAAGTGATA 410
QY 281 atatttcataatcccaagtagatttgataagcaaatctgttt-tatggaggaataattcattg 339
Db 409 TATAATTTAATTTATATATAAATTTAATTTACTTTCATTGATATATATAATTTAATTAATG 350
QY 340 gatttcacgaactcttaataagaattgatagaccacaaagaaatttagtataataacaa 399
Db 349 TACCTTTTCATAAATATTTATTTTATTAGTCTAGTAATATTTCTATTAAATAGTACCCCT 290
QY 400 cgagaaagtaataaaataataacttaactgaatgataacggaagatttgaatgaatagatct 459
Db 289 TTAATTTGGATATTACTACCTACTAAATATTTACCTA-ATAATATATTATTAAAGTAATCT 231
QY 460 atacttcccccccaacgctggtgcgtaaaaaagaaagtttccgataagtttaagtgatatt 519
Db 230 AAATCTAATAAATTTATTATCTFAAAGTATATAAATTAATTAATCTTTTTTATTATTATTT 171
QY 520 agagatttagtaactgttgaaagattatgattttgcaataaagagagcctctggctgatttc 579
Db 170 AAATTTATTATTAAATTTAGTAAATTTATTTATTTTATTATTAAACATAAATTTTGTGATA 111
QY 580 aaaaatcggttactcaataaagtagtacttttacagtagatgatttaacaga 626
Db 110 AATATATCATTTAATAATGGTAATTTATTAAATAATTTATCTTTAATGA 64
```

```
RESULT 4
US-08-998-416-287/c
; Sequence 287, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippsen, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jurgen
; APPLICANT: Knechtle, Philipp
; APPLICANT: Rebschung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPHII
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6239264artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: No. 6239264th Carolina
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,416
FILING DATE: 24-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 287:
SEQUENCE CHARACTERISTICS:
LENGTH: 860 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE: PAG1240UP
ORGANISM: US-08-998-416-287

Query Match 5.3%; Score 50; DB 4; Length 860;

Best Local Similarity 45.5%; Pred. No. 0.012;
Matches 252; Conservative 0; Mismatches 300; Indels 2; Gaps 2;

QY 74 taatcaaacgcttactgattgttgatcacaattgctcattgataatccaagtaggtg 133
DB 616 TAAGGAAATCGGATATTTTGCACTTAATTTTATTTTAAATTTGTTATTTATCTA 557
QY 134 atttaagcaattcttaacagataattcagttgtgagataatagaataaaactctgctta 193
DB 556 TTTAAACATAAAACATTTTAAATGTATATAAATAAATAAGAAATTACTTATAGATATT 497
QY 194 atgaagaaaatttggtttagcatcaagtttgacaacggtgaaaattcttaaggag 253
DB 496 ATTAATAGTATTTAATTTAATTTTAAATTAATATACCATTTTATTAATAAATAGAT 437
QY 254 aatatatttttagagtgatgctgatgatattctcatatccaagtagatttgataagcaaa 313
DB 436 TATTAGTTTATATATTAACTGATATATATTTAAATTTATATAAATTTAATTTAC 377
QY 314 ttctgttt-tatggagaaaattcattggtttctcagcaactctaatagaattgatagac 372
DB 376 TTCATTGATATATATAATTTAAATGTACCTTTTCATATAATTTATTTTATTAGTCTAG 317
QY 373 caaaaagaaaatttagtatataaaacacgagaaagataataataataacttaactaatgat 432
DB 316 TAAATTTCTATTATTAGTCTACCTTTAATTTGGATATTACTACTATAAATTTTAC 257
QY 433 atacggaagatgttattgaatagatctatacttgcaccaccaacggtggtgcgtaaaaag 492
DB 256 CTA-ATAATATATTATAGATATACTTAATCTAATAATTTTATTATCTAAGATATAA 198
QY 493 aaagtttctogaagttaattgggatagagatttagtacctgttgaagattatgatttt 552
DB 197 TTAATTAATCTTTTATTATTATTAAATTTATTATTAAATTTAGTAAATTTATATTATT 138
QY 553 gcaataagaggagctcgtgatttcaaaatcgcgttactcaataaagtaacttttaag 612
DB 137 ATTTTATTAACATAATTTTTTGTATAATAATATATCATTTATAAATCGTAAATTTATAA 78
QY 613 tatagattaaacga 626
DB 77 ATTATCTTTAATGA 64

RESULT 5

US-08-213-419B-3
; Sequence 3, Application US/08213419B
; Patent No. 6333406
; GENERAL INFORMATION:
; APPLICANT: Inselburg, J. et al.
; TITLE OF INVENTION: GENE ENCODING PROTEIN ANTIGENS OF PLASMODIUM FALCIPARUM
; FILE REFERENCE: JII-002CNC
; CURRENT APPLICATION NUMBER: US/08/213,419B
; CURRENT FILING DATE: 1994-03-14
; PRIOR APPLICATION NUMBER: US 07/870,506
; PRIOR FILING DATE: 1992-04-17
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 6124
; TYPE: DNA
; ORGANISM: Plasmodium falciparum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2407)..(2439)
; NAME/KEY: CDS
; LOCATION: (2598)..(3404)
; NAME/KEY: CDS
; LOCATION: (3580)..(3720)
; NAME/KEY: CDS
; LOCATION: (3850)..(5835)
US-08-213-419B-3

Query Match 5.3%; Score 50; DB 4; Length 6124;
Best Local Similarity 44.3%; Pred. No. 0.016;
Matches 203; Conservative 0; Mismatches 255; Indels 0; Gaps 0;

QY 298 agatttgataagcaaatctggtttatggaggaaaattcattggatttctcagcaactcta 357
DB 2169 aaattcaataaatttaataataataataattattattataattataatatataata 2228
QY 358 atagaattgatagacaaaagaaaatttagtatataaaacaacagaaaagtaataaata 417
DB 2229 ataaatttttttaatttaaaatttaatttagattgtccaaaaaaataaaaaataaa 2288
QY 418 tacttaactatgatatacggagatgttattgaatagatctatcttgcaccaccaacg 477
DB 2289 atatatataattataaaatacatataattataacataaagaaaataataatacaac 2348
QY 478 tgggtcgtaaaaaagaaagtttctgataagtttaattgggatatagagatttagacctgt 537
DB 2349 atattcaaaaaaattaaagttcttaaaattattataataaacttaattcatatcaaa 2408
QY 538 gaagatttatgatttgcataaagaggctcgtgctgatttcaaaatcggttactcaat 597
DB 2409 gaagtcataatacttctggttttttcataatcgtgaagaatgaaaaaataaaaaa 2468
QY 598 aaagtacttttaccagtatagattaaacgagaaatggaatatcacacccaataagtttaag 657
DB 2469 aaagaaaaaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaaag 2528
QY 658 caatatatttactcgtctattttacaagattttttataaagaaaaatcttatattgata 717
DB 2529 ataatattataataaataataataattttttacgcatacacacaacatttgcattat 2588
QY 718 acaaaaaattactaatttcaagagtatgataaa 755
DB 2589 tttttttttaggtggtatatttaacaaaaatgttataa 2626

RESULT 6

US-08-232-463-14/C
; Sequence 14, Application US/08232463
; Patent No. 5670367

GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHIEFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZgpt-F1s
US-08-232-463-14

Query Match 5.3%; Score 49.4; DB 1; Length 7218;
Best Local Similarity 9.1%; Pred. No. 0.022;
Matches 38; Conservative 200; Mismatches 181; Indels 0; Gaps 0;
Qy 123 aagtagaggtgatttaagcaattcttaacagaaatttcagttgtagataataagataaa 182
Db RRR 1343
Qy 183 aatcttgcttaagaaataattggttagcatcaagttgaacaaagcggtaaaat 242
Db RRR 1283
Qy 243 ttctaaggagaaataatttttagaagtgatgctgagataatttcataccaagtagatt 302
Db RRR 1223
Qy 303 tgaagaacaaattcttttagagagaaatttcattgattctcagacaactcaataga 362
Db RRR 1163
Qy 363 attgatgaccaaagaaatttagtatataaacaacagagaaagtaataaaatactt 422
Db RRR 1103
Qy 423 aactaatgataacggaagatgttatgaatagatctatacttcccaacccaacggtg 482
Db RRR 1043

Qy 483 cgtaaaaaagaaagtttttcgataaagtttaattggtgatatagagatttagtacctgttgag 541
Db 1042 CCAAGCTCGGAATTAATTTCTGTGAGCGTATGCGCAACGAAGAAAAATAGTTATAGTAG 984
RESULT 7
US-08-257-073-12
Sequence 12, Application US/08257073
Patent No. 5766597
GENERAL INFORMATION:
APPLICANT: Paoletti, Enzo
APPLICANT: de Taisne, Charles
APPLICANT: Tine, John A.
TITLE OF INVENTION: MALARIA RECOMBINANT POXVIRUS VACCINE
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Curtis, Morris & Safford, P.C.
STREET: 530 Fifth Avenue, 25th Floor
CITY: New York
STATE: New York
COUNTRY: UNITED STATES OF AMERICA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/257,073
FILING DATE: 09-JUN-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/075,783
FILING DATE: 11-JUN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/852,305
FILING DATE: 18-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/672,183
FILING DATE: 20-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Frommer, William S.
REGISTRATION NUMBER: 25,506
REFERENCE/DOCKET NUMBER: 454310-2570
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 840-3333
TELEFAX: (212) 840-0712
TELEX: 425066 CURTMS
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 1882 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-257-073-12

Query Match 5.2%; Score 48.8; DB 1; Length 1882;
Best Local Similarity 47.4%; Pred. No. 0.024;
Matches 146; Conservative 0; Mismatches 162; Indels 0; Gaps 0;
Qy 106 atttcattgataccaagtagagggtgatttaagcaattcttaacagaaatttcagtt 165
Db 466 ATAATATTGAAATCAATACTACTTTTATACACCGGTAGCTACCGGAAATCAATAT 525
Qy 166 gtagataataagaataaaaaattctgcttaataagaagaaatttggttttagcatcaagtttg 225
Db 526 TTAAGAATGAGGTTTGTCTTCTCCACAGAACCTCTTATGTCACCAATGACATTA 585
Qy 226 aacaaagcggtaaaattcttaaggagagaatatatttttagaagtgatgagattt 285
Db 586 GATGAAATGAGACATTTTATATAAAGATAAATAATATGTAATAAATTTAGATGAATTGACT 645


```

Query Match      5.2%; Score 48.6; DB 4; Length 615;
Best Local Similarity 49.2%; Pred. No. 0.022;
Matches 157; Conservative 0; Mismatches 159; Indels 3; Gaps 1;

Qy 576 ttccaaatcggtctactcaataaagctactctttacagtagatttaacagagaatggaat 635
Db 608 TTATAATATTTTATTAATAATTAATAATGATAAATAATTAATAA---ATAATTAATT 552

Qy 636 atcacaaaccaataagtttaagcaataatatttactcagctattttcaagaatttttataa 695
Db 551 TAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 492

Qy 696 agaaaaatcttatttgatcacaaaaattactaaattactttcgaagtagtgataaa 755
Db 491 ATAGAATAATTAAGCTTAATAATTAATAATAATAATAATAATAATAATAATAATAATA 432

Qy 756 gaaacgtctactcagcagcgtctctctaaatttttgagctaaattctacccttagtat 815
Db 431 AAATCAACATAATATTTATATAAATAATAGATATATATAAATAAATAATAATAATAATTT 372

Qy 816 tactattagaaactatatttcttatttatttactttaaagctcccttggttagag 875
Db 371 AAATAAATAATAATCTTATATAATAATAATAATAATAATAATAATAATAATAATAATAA 312

Qy 876 gttattaataaatgatatt 894
Db 311 TATATTAATTATTCATAAT 293.

RESULT 10
US-08-961-083-195
; Sequence 195, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 195:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 844 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-961-083-195

```

```

Query Match      5.1%; Score 47.6; DB 3; Length 844;
Best Local Similarity 50.4%; Pred. No. 0.039;
Matches 116; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

Qy 103 ataattgcttgataatcccaagtagagtgatttaaagcaattcttaacagaatattcoa 162
Db 44 ATTATTTTGTAGTTGACGGTTCTACGGATAATTCTGGGAAATTTGTGATGCTTTTATG 103

Qy 163 gttgataataaataaaatcttcttaataagaagaaatattggttttagcatcaagt 222
Db 104 ATGCAAGATAATCGTGTGCGAGTATTGTCATCAAGAAATAAGGGGGGCGACACAAGCT 163

Qy 223 ttgacaacagcgtgaaatcttctaaagggagaataatttttagaatgtagctgatgat 282
Db 164 AAAATATGGGATAGTAGTGACCTAAGGAGAGTACATCAGCATTTGTTGATTCAGATGAT 223

Qy 283 atttcataatcccaagtagatttgataagcaaatctgttttagggagaaaa 332
Db 224 ATCGTAAAGAAATAATGATTGAAACTCTTTATATCAGCAAGTCCCAAGAAAA 273

RESULT 11
US-08-998-416-288/c
; Sequence 288, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippsen, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jürgen
; APPLICANT: Knechtle, Philipp
; APPLICANT: Rebeschung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPHII
; NUMBER OF SEQUENCES: 1152
; NUMBER OF INVENTION: AND USES THEREOF
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6239264artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: No. 6239264th Carolina
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,416
; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 288:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 837 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: PAG124IRP
; US-08-998-416-288

```


TELEFAX: (206)682-6031
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3701 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-990-571-10

Query Match 4.9%; Score 46; DB 4; Length 3701;
Best Local Similarity 46.3%; Pred. No. 0.11;
Matches 151; Conservative 0; Mismatches 175; Indels 0; Gaps 0;

Qy 593 tcaataaagctactttacagtagatgaacgagaatggaatatcacaaaccaataagt 652
Db 581 TAATTATAATCAATTATTATATATATATATATGGATATGATATATATCATTAATGATA 522
Qy 653 ttaagcaatatatttactcagctattttacaagattttttaagaaaaatcttatattg 712
Db 521 CAATAACAATATGTAATTATATAAATATCAATAATAATCATATATAATAAATTAATATA 462
Qy 713 atatacaaaaaattactaattactttcaagagtagtgataaagaacgcctatactcagc 772
Db 461 ATATTGCAATTATTAATTATATATCTGTTCTTAATTTATATCATTAACAATACAAAACATA 402
Qy 773 aagagctctctaaatatatttgagctaaatctacccctagattactattagaaaactat 832
Db 401 TATACITTTTATAATTATAACACATGAATATAATTAATGATATTAACAATGCCTATG 342
Qy 833 atatttggttattttactttaaagtcctccctggtagaggttattataaataagata 892
Db 341 ATGATTATTATCATCTTAAATTCGTCAATTAATCCCTTAAGATCTTATCTAATTTAT 282
Qy 893 ttaatatatttagtactgaaattgttt 918
Db 281 TTAATATTCTATATCTGATATATTTT 256

RESULT 14

US-08-723-142A-10/G
; Sequence 10, Application US/08723142A
; Patent No. 6306396
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond
; APPLICANT: Sleath, Paul R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
; AND TREATMENT OF B. MICROTI INFECTION
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/723,142A
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Makl, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.426
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900

TELEFAX: (206)682-6031
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3701 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-723-142A-10

Query Match 4.9%; Score 46; DB 4; Length 3701;
Best Local Similarity 46.3%; Pred. No. 0.11;
Matches 151; Conservative 0; Mismatches 175; Indels 0; Gaps 0;

Qy 593 tcaataaagctactttacagtagatgaacgagaatggaatatcacaaaccaataagt 652
Db 581 TAATTATAATCAATTATTATATATATATATGGATATGATATATATCATTAATGATA 522
Qy 653 ttaagcaatatatttactcagctattttacaagattttttaagaaaaatcttatattg 712
Db 521 CAATAACAATATGTAATTATATAAATATCAATAATAATCATATATAATAAATTAATATA 462
Qy 713 atatacaaaaaattactaattactttcaagagtagtgataaagaacgcctatactcagc 772
Db 461 ATATTGCAATTATTAATTATATATCTGTTCTTAATTTATATCATTAACAATACAAAACATA 402
Qy 773 aagagctctctaaatatatttgagctaaatctacccctagattactattagaaaactat 832
Db 401 TATACITTTTATAATTATAACACATGAATATAATTAATGATATTAACAATGCCTATG 342
Qy 833 atatttggttattttactttaaagtcctccctggtagaggttattataaataagata 892
Db 341 ATGATTATTATCATCTTAAATTCGTCAATTAATCCCTTAAGATCTTATCTAATTTAT 282
Qy 893 ttaatatatttagtactgaaattgttt 918
Db 281 TTAATATTCTATATCTGATATATTTT 256

RESULT 15

US-08-998-416-1137/C
; Sequence 1137, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippsen, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jürgen
; APPLICANT: Knechtie, Philipp
; APPLICANT: Reibischung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
; AND USES THEREOF
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NO. 6239264artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: No. 6239264th Carolina
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,416
; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996

```

; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 1137:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 636 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: PAG1692RP
; US-08-998-416-1137

```

```

Query Match      4.9%; Score 45.6; DB 4; Length 636;
Best Local Similarity 50.6%; Pred. No. 0.099;
Matches 162; Conservative 0; Mismatches 154; Indels 4; Gaps 2;

QY 576 ttctaaatcggcttactcaataagtgacttttacagtgatagattaaacga-gaatggaa 634
   || || || || || || || || || || || || || || || || || || || || ||
Db 609 TTATAATATTTTATTAAATTAATTAATAATGATAAAATATTAAATTAATTAATTTAA 550

QY 635 tatcacaacccaataagtttaagcaatatatttactcagctattttacagattttata 694
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 549 TTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 490

QY 695 agaaaaatcttatgtatgatcacaaaaattactaattactttcaagagtgatgataa 754
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 489 AGAAATTAAAGTTAAATTAATTTTAATAATAATCTTTATAAAAGATTAAATAATAATA 430

QY 755 agaaacgctatactcgaagagctctcctaaatattttgagctaaatactacccctagta 814
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 429 ATCACATAATAATTTATAAAATAGATAT--TATAATAAAATAATAATAATTTACATATT 373

QY 815 tiactattagaaactatatttggttatttatacttttaagctcctccctggtagga 874
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 372 TAAATAAATTAATCTTTATAATAATAATAATAATAATAATAATAATAATAATAATA 313

QY 875 ggttattaataaatgatatt 894
   ||||| | | | | |
Db 312 ATATATTAATTATTGATAAT 293

```

Search completed: August 2, 2002, 13:28:36
Job time: 9682 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 31, 2002, 18:56:25 ; Search time 13.52 Seconds
(without alignments)
896.392 Million cell updates/sec

Title: US-09-900-038A-1
Perfect score: 1590
Sequence: 1 MNYGIIMSVYNEPLNYRDS.....LINDINILVLKFGKQSD 313

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	437.5	27.5	290	1 YAS7_METJA	Q58457 methanococ
2	284	16.6	323	1 YF78_HAEIN	Q57287 haemophilus
3	206.5	13.0	267	1 YG95_HAEIN	Q48215 haemophilus
4	199.5	12.5	266	1 AMSE_ERWAM	Q46635 erwania amy
5	196.5	12.4	294	1 YG96_HAEIN	Q48214 haemophilus
6	192	12.1	900	1 GGAB_BACSU	Q46918 bacillus su
7	189	11.9	250	1 Y868_HAEIN	Q57022 haemophilus
8	182.5	11.5	344	1 YIBD_ECOLI	P11290 escherichia
9	175	11.0	256	1 SP5A_BACSU	P39621 bacillus su
10	161	10.1	279	1 WCAA_ECOLI	P77414 escherichia
11	158.5	10.0	909	1 Y4GI_RHISN	P55465 rhizobium s
12	145	9.1	268	1 WDFE_BACSU	P39614 bacillus su
13	144.5	9.1	322	1 Y86E_ANASP	P22639 anabaena sp
14	140	8.8	323	1 YKCC_BACSU	Q34319 bacillus su
15	129	8.1	301	1 AMSE_ERWAM	Q46632 erwania amy
16	125	7.9	260	1 DPM1_MOUSE	Q70152 mus musculus
17	123	7.7	266	1 DPM1_CRIGR	Q9wu83 cricetulus
18	122.5	7.7	346	1 YF20_MYCTU	Q50587 mycobacteri
19	119	7.5	260	1 DPM1_HUMAN	O60762 homo sapien
20	119	7.5	581	1 Y208_BORBU	O51226 borrelia bu
21	119	7.5	1197	1 DPON_PODAN	Q01529 podospora a
22	118	7.4	446	1 GGAA_BACSU	P46917 bacillus su
23	116.5	7.3	979	1 P115_MYCHR	P41508 mycoplasma
24	116	7.3	1805	1 HMW2_MVCGE	P47460 mycoplasma
25	114.5	7.2	236	1 DPM1_SCHPO	O14466 schizosacch
26	113.5	7.1	993	1 NISE_LACLA	P20103 lactococcus
27	113	7.1	569	1 CISP_PLAFA	P25805 plasmodium
28	112	7.0	419	1 HASA_STRPY	Q54865 streptococ
29	110.5	6.9	997	1 T257_ECOLI	P25239 escherichia
30	110	6.9	2273	1 HFAL_YEAST	P32874 saccharomyc
31	109.5	6.9	653	1 MTSI_STRSA	P29347 streptococ
32	109	6.7	970	1 Y087_BUCAI	P57189 buchnera ap
33	107	6.7	1169	1 EX5E_BORBU	O51578 borrelia bu

ALIGNMENTS

RESULT 1
YA57_METJA STANDARD; PRT; 290 AA.
AC Q58457;
DT 15-JUL-1998 (Rel. 36, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative glycosyl transferase MJ1057 (BC 2.-.-.-).
GN MJ1057.

OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_TaxID=2190;
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Usterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
jannaschii.";
RL Science 273:1058-1073(1996).

CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U67549; AAB99061.1; -.
DR HSSP; P39621; 1QG8.
DR TIGR; MJ1057; -.
DR InterPro; IPR001173; Glycos.transf_2.
DR Pfam; PF00535; Glycos.Transf_2; 1.
KW Hypothetical protein; Transferase; Complete proteome.
SQ SEQUENCE 290 AA; 35099 MW; 3F6A1B221C420D74 CRC64;

Query Match 27.5%; Score 437.5; DB 1; Length 290;

Best Local Similarity 35.2%; Pred. No. 1.7e-22;

Matches 99; Conservative 58; Mismatches 101; Indels 23; Gaps 8;

QY 4 SLIMSVYNEPLNYRDSVESILNQTLDTFEPIIVIDNPGRDLKQFLTEYSWPNRIKIL 63

Db 11 SVVWATNEPEKYLUKESIESIXNXTKDFEPIIVLDNPNKKAEITEKEYQOKRIRIFI 70

QY 64 LNEENIGLASSLNKAVKISKGEYIFRMDADDISYPSRFDKQIRFMEEN-SLDFSATLIEL 122

```
Db 71 KNERLGRGASRKNKAVNTARGKYIAILDADDIALPKRLEKOFKYMENNRDIDLLESWYF 130
QY 123 IDQGNLVYKQRESNKIYLTNDIRKMLNRSILAHPTWCVKKKVFDKLMGYRDLVPVEDY 182
Db 131 IDENGIL-KEPKPEK-YKFEIKKYFFKEHLTVHPMMVSKSLKLLKYDEKLLRSQDY 188
QY 183 DPAIGALADFKIGLNLKVLQYRL--NENGISQTNKFKQYIYSAILQDFYKEKSY---- 236
Db 189 DFWIRCIANDYKFDIEEFLLKYRIPNRDNYLSRIKKQKLYSYT-LKTHWKNKKHFCNN 247
QY 237 IDITKITNYFQBYVVIKKRYTQOELSYPKLKSTPSITIRKL 277
Db 248 VYFVKV--FFYSLVV-----YLFIVLTPTFIILKIL 275

RESULT 2
YF78_HAEIN STANDARD: PRT; 323 AA.
AC Q57287; O05077;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative glycosyl transferase H11578 (EC 2.-.-.-).
GN H11578.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RA "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd.";
RL Science 269:496-512(1995).
CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U32832; AAC23227.1; -.
CC TIGR; H11578; -.
CC InterPro; IPR001173; Glycos_transf_2.
CC Pfam; PF00535; Glycos_transf_2.1.
CC KW Hypothetical protein; transferase; Glycosyltransferase;
CC Complete proteome.
SQ SEQUENCE 323 AA; 37680 MW; 7CBC2681039AB5B4 CRC64;

Query Match 16.6%; Score 264; DB 1; Length 323;
Best Local Similarity 29.0%; Pred. No. 6.5e-11;
Matches 93; Conservative 62; Mismatches 114; Indels 52; Gaps 13;

QY 4 STMSVYNEPLNYVRDSVESILNQTLDFEFIVINDPNRSGDLKQFLTEYSVVDNRKIL 63
Db 8 SVIVCAYNAE-QYIDESISSLIQYVENLE-IVINDGSDTLTSLHLEISIKLDKRKII 65
QY 64 LNEENIGLASSLUNKAVKISKGEYIFRMDADDISYSPRFDKQIRFMEENS-LDFSATLIEL 122
```

```
Db 66 SNKYNLGFINSLNIGLGCFSKGYFARMDDADDIAKPSWIEKIVTYLEKNDHITAMGSYLEI 125
QY 123 IDQK-NLVYKQRESNKIY----LTNDIRKMLNRSILAHPTWCVKKKVF--DKLMGYRD 175
Db 126 IVECEGIIQSGYKTDIWKPNPLHNDICEAMLFNPHNNTWIMRWANVYRHKILFNKD 185
QY 176 LVPVEDYDF----AIRGALADFKIGLNLKVLQYRLNENGISQTNKFKQYIYSAILQ---- 228
Db 186 YPAEDYKFWSEVSRGLGLANYP-----EALVKYRLHGNOTSSVYNHEQNETAKKIKREN 240
QY 229 -DFYKEKSYDITKITN-----YFQBYVVIKKRYTQOEL---SKY--- 263
Db 241 ITYLNKIGIDIKVINSVLSLEIYHVDKSNKVLKSILYEMYSLDKYTTITSLHFIKYHL 300
QY 264 --FELKSTPSIT---IRKLYI 279
Db 301 ELFDLQNLKIITKIFIRKINV 321

RESULT 3
YG95_HAEIN STANDARD: PRT; 267 AA.
AC Q48215; O05081;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Putative glycosyl transferase H11695 (EC 2.-.-.-).
GN H11695.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A2;
RA McLaughlin R., Abu Kwaik Y., Young R., Spinola S., Apicella M.;
RT "Characterization and sequence of the lsg locus from Haemophilus
RT influenzae.";
RL Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RA "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd.";
RL Science 269:496-512(1995).
CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M94855; AAA24983.1; -.
CC DR EMBL; U32842; AAC23341.1; -.
CC TIGR; H11695; -.
CC InterPro; IPR001173; Glycos_transf_2.
CC Pfam; PF00535; Glycos_transf_2.1.
CC KW Hypothetical protein; transferase; Glycosyltransferase;
CC Complete proteome.
```

FT CONFLICT 26 26 V -> G (IN REF. 1).
 FT CONFLICT 46 46 D -> E (IN REF. 1).
 FT CONFLICT 49 49 F -> S (IN REF. 1).
 SQ SEQUENCE 267 AA; 30770 MW; A2F1A0532737D8C3 CRC64;

Query Match
 Best Local Similarity 13.0%; Score 206.5; DB 1; Length 267;
 Matches 70; Conservative 57; Mismatches 130; Indels 25; Gaps 8;

QY 1 MNYSIINSYV-NEPLNYYRDSVESILNQTLDTEFEIIVIDNPSRGLDKQFLTEYSVVDNR 59
 Db 1 MKFSVLMSLIKENPOQLFCFESLVAQTRQADEIVLVGCVTPDPEFVYTFEE-TKLP 59
 QY 60 IKILLNEENIGLASSLNKAVKISKGEYIFRMDADDISYPSRFDKQIRFMEEN--SLDPSA 117
 Db 60 LKLVLPQNLGLKALNEGILHCDYDWMFRMDTDDICVDPDFEKQAVAFIQHPESIIFGG 119
 QY 118 TLIELIDOKGNLV-YKQRESNKIYLTNDIRKMLNRSILAHPTWCVKKKVFDKLMGYRDL 176
 Db 120 QIAEFGRNVDIVAYRNVPTS---AQEIIKFTOKRCPFNHMTVAYQKSAVINGCGYEDL 175
 QY 177 VPVEDYDFAIRGALADFKIGLLNKVLQYRLNENGISQTNKFKQIYISAILQDFYKESY 236
 Db 176 --QEDYVLTWIKLVAQGYMANLPDILYVYRGNGWVSRRRGVNQ---AKAEWRLEKLYR 230
 QY 237 IDITKITNYFOEYVKKRYTQQLSKYFELKSTPSIIRKLY 278
 Db 231 LGTQGLSLGFTFAIR-----FGSRLPTSLKKLY 261

RESULT 4
 AMSE_ERWAM STANDARD; PRT; 266 AA.
 AC Q46635;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Amylovoran biosynthesis glycosyl transferase amse (EC 2.-.-.-).
 GN AMSE.
 OS Erwinia amylovora.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Erwinia.
 OX NCBI_TaxID=552;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=EAL/79;
 RX MEDLINE=95319333; PubMed=7596293;
 RA Bugert P., Geider K.;
 RT "Molecular analysis of the amse operon required for exopolysaccharide
 synthesis of Erwinia amylovora.";
 RL Mol. Microbiol. 15:917-933(1995).
 CC -1- FUNCTION: INVOLVED IN THE BIOSYNTHESIS OF AMYLOVORAN WHICH
 CC FUNCTIONS AS A VIRULENCE FACTOR.
 CC -1- PATHWAY: EXOPOLYSACCHARIDE BIOSYNTHESIS.
 CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@lsb-sib.ch).
 CC -----
 DR EMBL; X77921; CAA54886.1; -;
 DR InterPro; IPR001173; Glycos_transf_2.
 DR Pfam; PF00535; Glycos_transf_2; 1.
 KW Exopolysaccharide synthesis; Transferase; Glycosyltransferase.
 SQ SEQUENCE 266 AA; 30748 MW; 163268A4210EB47B CRC64;

Query Match
 12.5%; Score 199.5; DB 1; Length 266;

Best Local Similarity 28.4%; Pred. No. 9.9e-07;
 Matches 73; Conservative 48; Mismatches 107; Indels 29; Gaps 12;

QY 3 YSIIMSIVN--EPLNYYRDSVESILNQTLDTEFEIIVIDNPSRGLDKQFLTEYSVVDNR 60
 Db 2 FSVLISLNKPKEN-LEQCLESQHQTLDNADEIVLVYDGPVBSLSKAVATRW---NLL 57
 QY 61 KILL--NEENIGLASSLNKAVKISKGEYIFRMDADDISYPSRFDKQIRFMEEN--SLDPSA 117
 Db 58 PLVIVPLEKNLGLKALNAGLERCTHNVARMDTDDICLPERFEKQISYMESHPEVVLGS 117
 QY 118 TLIELIDOKGNLV-YKQRESNKIYLTNDIRKMLNRSILAHPTWCVKKKVFDKLMGYRDL 172
 Db 118 AAVIEFDEHG---KERLRLPLSNNDIHEFARMKNPNH--MCV---VFRKDKVISAGS 168
 QY 173 YRDVLPVEDYDFAIRGALADFKIGLLNKVLQYRLNENGISQTNKFKQIYISAILQDFYK 232
 Db 169 YQHLYLMEDYNLMURIMSLGHPVANLPDVLKVR---GSDMVNKRNGWYIKSEVQLYR 225
 QY 233 EKSVIDIT--KITNYF 246
 Db 226 LKALKQTGFIRGTLXF 242

RESULT 5
 YG96_HAEIN STANDARD; PRT; 294 AA.
 AC Q48214; Q05082;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Putative glycosyl transferase H11696 (EC 2.-.-.-).
 GN H11696.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Haemophilus.
 OX NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A2;
 RA McLaughlin R., Abu Kwaik Y., Young R., Spinola S., Apicella M.;
 RT "Characterization and sequence of the lsg locus from Haemophilus
 influenzae.";
 RL Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RD / KW20 / ATCC 51907;
 RX MEDLINE=95350630; PubMed=7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
 RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 RA Venter J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
 Rd.";
 RL Science 269:496-512(1995).
 CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@lsb-sib.ch).
 CC -----
 DR EMBL; M94855; AAA24982.1; -;
 DR EMBL; U32842; AAC23342.1; -;
 DR TIGR; H11696; -;

DR InterPro: IPR001173; Glycos_transf_2.
DR Pfam: PF00535; Glycos_transf_2; 1.
KW Hypothetical protein; Transferase; Glycosyltransferase;
KW Complete proteome.
FT CONFLICT 38 C -> Y (IN REF. 1).
FT CONFLICT 48 S -> R (IN REF. 1).
FT CONFLICT 70 V -> I (IN REF. 1).
FT CONFLICT 74 T -> A (IN REF. 1).
FT CONFLICT 88 C -> R (IN REF. 1).
FT CONFLICT 97 V -> I (IN REF. 1).
FT CONFLICT 106 N -> D (IN REF. 1).
FT CONFLICT 152 I -> V (IN REF. 1).
FT CONFLICT 185 P -> S (IN REF. 1).
SQ SEQUENCE 294 AA; 33646 MW; 8330F081BFF7A18B CRC64;

Query Match 12.4%; Score 196.5; DB 1; Length 294;
Best Local Similarity 25.1%; Pred. No. 1.8e-06;
Matches 81; Conservative 60; Mismatches 115; Indels 67; Gaps 15;
QY 4 SLIMSVYNEPLNVRDVSVESILNQTLTDFEIIIVDNPFRGDLKQFLTEYSVVDNR--- 59
DB 3 STIIVSYNRKAE-VPALLESLTQQTSSNFE-VLIIVDCSK-----ERVVVEQSYSEFP 52
QY 60 IKILNEENIGLASSLNKAVKISKGEYIFRMDADDISYPSRFDKQIRFMEEN---SLDPS 116
DB 53 VTVIRNETNQGAESRNVGARTSGDWLLFLDDDDCFWPEKCEKVLQVTEQNPINFIYH 112
QY 117 APLIELDQKGNLVYKQESNKIYLTNDIRKMLNRSILAHPTWCVKKVDKLMGYRD- 175
DB 113 PAKCEWNEGFTYVTOPIEQEI-----STERILLANKIGGMPWIAIKKEMFLUKIGLSTA 168
QY 176 LVPVEDYDFAIRGALADFKIGLLNKVLQ-----YRLNE-----NGISQ 214
DB 169 LRSLEDYDFLLK-----LLQEPSFTYKINEPLTYCTFHTKRSVSDITYN 214
QY 215 TNKFKQYYSAILQDPYEKSYIDITKTNTYQEVYIKRKYTOQLSKYFEL-KSTPSIT 273
DB 215 TOKAIDYIRHVYKTVQEARNF-DIN--ASYILAYPHIMNLNRKAAYKVFDFIKTKSI- 270
QY 274 IRKLYICLYLXPKSLVRLRLIN 296
DB 271 --KQFIITLVILISP---KLAIN 288

RESULT 6
GGAB_BACSU STANDARD; PRT; 900 AA.
AC P46918;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Minor teichoic acids biosynthesis protein ggab.
GN GGAB.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Freymond P., Karamata D.;
RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: INVOLVED IN THE BIOSYNTHESIS OF GALACTOSAMINE-CONTAINING
CC MINOR TEICHOIC ACIDS.
CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC or send an email to license@isb-sib.ch.
DR EMBL: U13979; AAA73513.1; .
DR EMBL: Z99122; CAB15585.1; .
DR Subtilist; BG11192; ggab.
DR InterPro: IPR001173; Glycos_transf_2.
DR Pfam: PF00535; Glycos_transf_2; 1.
KW Transferase; Glycosyltransferase; Complete proteome.
SQ SEQUENCE 900 AA; 107154 MW; FA66495488C2C62F CRC64;

Query Match 12.1%; Score 192; DB 1; Length 900;
Best Local Similarity 23.7%; Pred. No. 1.4e-05;
Matches 88; Conservative 69; Mismatches 130; Indels 84; Gaps 19;
QY 2 NYSIISVYNEPLNVRDVSVESILNQTLTDFEIIIV-----DNPSRGDL--KQFLTEYSV 55
DB 9 DFSVIMPIYNVEL-YLTAIESIINQTI-GFENIQILVNDSDPKSEIICKEYAQKY-- 64
QY 56 VDNRIKILLNEENIGLASSLNKAVKISKGEYIFRMDADDISYPSRFDKQIRFMEENSLDF 115
DB 65 -PNNI-VYAKKQNGGVSSARNYGLKVAEGRYIQFLDPPDLVSEGTFFENVLNFDEHKNEI 122
QY 116 SATLIELDQKGNLVYKQESNKIYLTN-DIRKMLNRSILAHPTWC-----VKKKVFQK 169
DB 123 DIVAIPFFAEGR-TGEHNLNKFSTRILDVEKE--PHHILTH---CGSTFIKKDALKN 176
QY 170 LMGVYRDLVPVEDYDFAIRGALADFKIGLLNKVLQVRLNENGIS--QTNKFKQYIYSAIL 227
DB 177 IRFDENCKIGEDAKLVNLIISOKKYGVLKAEKYHYRVREDGSSAMQTAKANKNWFNHSL 236
QY 228 QDFYKESYIDITKTNTY-----FQEYVI-----KKRYTQ----- 257
DB 237 IFF--SKNLIDIIK--NHEQKIPFLQYVMVMDLKWKLKIDISETPLDENEYSEPLTLI 292
QY 258 QELSXYFE---LKSTPSITIRKLYICL-----YLYFKSPVLVRRL 293
DB 293 REVLSYIDDDVLIETKSVSHFYHALKIKHGENYSRYVYVERETEQDYLYLREGKIVSKL 352
QY 294 LINDINILVK 304
DB 353 SDQTLTIEILE 363

RESULT 7
Y868_HAEIN STANDARD; PRT; 250 AA.
AC Q57022; P96336;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative glycosyl transferase HI0868 (EC 2.-.-.-).
GN HI0868.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerkvliet A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd.";
RL Science 269:496-512(1995).

```
CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U32768; AAC22526.1; -.
DR HSP; P39621; IQG8.
DR TIGR; HI0868; -.
DR InterPro; IPR001173; Glycos_transf_2.
DR Pfam; PF00535; Glycos_transf_2; 1.
DR Hypothetical protein; Transferase; Glycosyltransferase;
KW Complete proteome.
SQ SEQUENCE 250 AA; 28915 MW; A5D8220129782E98 CRC64;

Query Match 11.9%; Score 189; DB 1; Length 250;
Best Local Similarity 27.08; Pred. No. 4.6e-06;
Matches 68; Conservative 53; Mismatches 107; Indels 24; Gaps 10;

Qy 4 SIIMSVYNEPLNYRDSVESILNLTDFEIIIVIDNPGRGLDKQFLTEYSVVDNRKIL 63
Dbb 7 SIIMPVNAEC-YNQGLTSLNQSQYQNI-LILIDGSDTKSIEIINNIIDKDKRVKLF 64
Qy 64 LNEENIGLASSLNKAVKISKGEYIFRDADDISYPSRFDKQIRWEENSLDFSATLIELI 123
Dbb 65 FTPTNQGPAARNGLEKAQGDYITFLDSDDFIANDKLEKQLNFMQLN-----HLV 115
Qy 124 DQKNLVYKQRESNKIYLTNDIRKM-LLNRSILAHPTWCVKKKYFDKLMGYRDLVPV 179
Dbb 116 MTHGNVAFCDLEGQIKLVTSKIDYITLQGNQKIMTVLVEREST-KULRFPN-IKH 173
Qy 180 EDYFAIRGALADPKIGLL--NKVLLQYRLNENGISOTNKFQYIYSAILQDFYKESYI 237
Dbb 174 EDYAFLL-DCLKEVKQSYLSHQASFFVRIGKVSVS-SNKPKSALWTF---NIVFKREKL 228
Qy 238 DITKITNYFQY 249
Dbb 229 GVKSIIYFILDY 240

RESULT 8
YIBD_ECOLI
ID YIBD_ECOLI STANDARD; PRT; 344 AA.
AC P11290; P22999;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative glycosyl transferase YibD (EC 2.-.-.-).
GN YIBD OR B3615.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OC NCBI_TaxID=562;
RX [1]
SEQUENCE FROM N.A.
STRAIN=K12 / MG1655;
MEDLINE=94316500; PubMed=8041620;
RA Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;
RT "Analysis of the Escherichia coli genome. V. DNA sequence of the
RT region from 76.0 to 81.5 minutes."
RL Nucleic Acids Res. 22:2576-2586(1994).
RN [2]
RN SEQUENCE OF 1-198 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=89174812; PubMed=2647748;
RA Aronson B.D., Somerville R.L., Epperly B.R., Dekker E.E.;
RT "The primary structure of Escherichia coli L-threonine
RT dehydrogenase."
RX MEDLINE=95020537; PubMed=7934828;

RL J. Biol. Chem. 264:5226-5232(1989).
CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U00039; AAB18592.1; -.
DR EMBL; AE000439; AAC76639.1; -.
DR EMBL; X06690; CAA29885.1; -.
DR PIR; A30268; Q3ECTH.
DR PIR; B33276; B33276.
DR EcoGene; EG11266; yibD.
DR InterPro; IPR001173; Glycos_transf_2.
DR Pfam; PF00535; Glycos_transf_2; 1.
DR Hypothetical protein; Transferase; Glycosyltransferase;
KW Complete proteome.
FT CONFLICT 1 26 MNSTNKLVSIIPLYNAGDDFTCME -> MRAMISALVWK
FT CONFLICT 77 77 R -> G (IN REF. 2).
SQ SEQUENCE 344 AA; 40524 MW; 9B5DABFE86D5457A CRC64;

Query Match 11.5%; Score 182.5; DB 1; Length 344;
Best Local Similarity 22.2%; Pred. No. 1.8e-05;
Matches 74; Conservative 67; Mismatches 117; Indels 75; Gaps 13;

Qy 4 SIIMSVYNEPLNYRDSVESILNLTDFEIIIVIDNPGRGLDKQFLTEYSVVDNRKIKI- 62
Dbb 9 SVIIPLYNAGDDF-RTCHESLITQTWALEIIINDGST-----DNSVEIA 53
Qy 63 -----LLNENIGLASSLNKAVKISKGEYIFRDADDISYPSRFDKQIRFMEEN 111
Dbb 54 KYAENYPHVRLHQANAGASVARNRGIEVATGKYVAFVDADDEVYPTMYETLMTMALED 113
Qy 112 SLDFSATLIELIDQGNLVYKQRESNKIYLTN-----DIRKMLNRSILAHPTW----- 160
Dbb 114 DLDVAOCNADWCFFRETGETWQSIPTDRLRSTGVLTGPDLNGLSSRRWTHVVMGVYRR 173
Qy 161 --CVKKV-FDKLMGYRDLVPVEDYDF-AIRGALADFKIGLNLKVLQYRLNENGIS--- 213
Dbb 174 DVIVKNNIKFIAGLHHQDIVWTTEPFNALRYTE-----QSLYKYLHNTSVSRLLH 226
Qy 214 -OTNKFQYIYSAILQDFYKESYIDITKI-----TNYFQEVYIKKRYTQ-----ELSKY 263
Dbb 227 RQGNKNLNY-----QRHYIKITRLLEKLNRYADKIMYEPHFQQTYYEALRVC 275
Qy 264 FELKSTPSITIRKLYICILYLFKSPLVRLIN 296
Dbb 276 HAVRKEPDILTRQRIA--EIFTSGMYKRLLTN 306

RESULT 9
SPSA_BACSU
ID SPSA_BACSU STANDARD; PRT; 256 AA.
AC P39621;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Spore coat polysaccharide biosynthesis protein spsA.
GN SPSA OR IPA-63D.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RX [1]
SEQUENCE FROM N.A.
RP STRAIN=168;
RC MEDLINE=95020537; PubMed=7934828;
```



```
Db 123 DYVQ--GEVYSOPASLPYKPSYRRLFYKRNIIGNOVFTWAWRKECLFD-----TE 175
Qy 176 LVPVEDYDFAIRGAL---ADPKILLNKLVLQYRLNENGI---SQTNKFQYIYSAILQD 229
Db 176 LKAQDIDIFLRVVEYGEWPVKVEATQIL---HINIGEMQITSSPKKFSGYFH----- 226
Qy 230 FY-KESYID-ITKITNYFOEYVIK-KRYTQOEL 260
Db 227 FYRKHKKDFDRASKYQLETLTYQIRNKRMTWRTL 260

RESULT 11
Y4GI_RHISN
ID Y4GI_RHISN STANDARD; PRT; 909 AA.
AC P55465;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 102.8 kDa protein Y4GI.
GN Y4GI.
OS Rhizobium sp. (strain NGR234).
OG Plasmid sym pNGR234a.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=394;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97305956; PubMed=9163424;
RA Frelberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,
RA Perret X.;
RT "Molecular basis of symbiosis between Rhizobium and legumes.";
RL Nature 387:394-401(1997).
CC -1- SIMILARITY: STRONG TO M.XANTHUS O-ANTIGEN BIOSYNTHESIS PROTEIN
CC REBC (NOT TO BE CONFUSED WITH ENTEROBACTERIAL REBC).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000074; AAB91683.1; -
DR HSSP; P39621; LQG8.
DR InterPro; IPR001173; Glycos_transf_2.
DR Pfam; PF00535; Glycos_transf_2; 2.
KW Hypothetical protein; Plasmid.
SQ SEQUENCE 909 AA; 102827 MW; 3985D69722F43962 CRC64;

Query Match 10.0%; Score 158.5; DB 1; Length 909;
Best Local Similarity 22.6%; Pred. No. 0.0023;
Matches 53; Conservative 51; Mismatches 102; Indels 29; Gaps 6;

Qy 4 SIIMSVYNEPLNYVRDSVESILNQTLDPEFIIVIDNPSRGDLKQFLTEYSVVDNRKIL 63
Db LVVVVYVNPDPALLVEMESVRAQSYANWELCLDDCSTDPEVGRVLRNRYAAQDPVRVV 428
Qy 64 LNEENIGLASSLNKAVISKGEYIFRMDADDISYPSRFDKQIRWEENSDFSATLIELI 123
Db 429 FREANGHMSQASNSAIEIARGAYIALDHDLDLP-----DALVLVVQVI 473
Qy 124 DQGNLYVKQRESNKIYLTNDIR-----KMLNRSILAHPTWCVKKKVFDKLM-----GY 173
Db 474 DAHPDAKIITYDEDKI-VEGTRCDAHFKPDWNRDLGYINGIYISHLGVDFDAALVREVGAF 532
Qy 174 RD-LVPVEDYDFAIR--GALADFKIGLLNKVLQYRLNENGISOTNKFQYIYA 225
Db 533 REGFEGAGDYDMLLRCTERVDQRQIHIIAKVLYSWRTPGSAASNRKAPYANEA 587

RESULT 12
```

```
YWDF_BACSU
ID YWDF_BACSU STANDARD; PRT; 268 AA.
AC P39614;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative glycosyl transferase ywdf (EC 2.-.-.-).
GN YWDF OR IPA-56D.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=95020537; PubMed=7934828;
RA Glaser P., Kunst F., Arnaud M., Coudart M.P., Gonzales W.,
RA Hullo M.F., Ionescu M., Lubochinsky B., Marcelino L., Moszer I.,
RA Presecan E., Santana M., Schneider E., Schweizer J., Vertes A.,
RA Rapoport G., Danchin A.;
RT "Bacillus subtilis genome project: cloning and sequencing of the 97
RT kb region from 325 degrees to 333 degrees.";
RL Mol. Microbiol. 10:371-384(1993).
CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X73124; CAAS1612.1; -
DR EMBL; Z99123; CAB15824.1; -
DR PIR; S39711; S39711.
DR Subtilist; BGI0602; ywdf.
DR InterPro; IPR001173; Glycos_transf_2.
DR Pfam; PF00535; Glycos_transf_2; 1.
KW Hypothetical protein; Transferase; Glycosyltransferase;
KW Complete proteome.
SQ SEQUENCE 268 AA; 30616 MW; DD8428F7016EC9B3 CRC64;

Query Match 9.1%; Score 145; DB 1; Length 268;
Best Local Similarity 25.1%; Pred. No. 0.0042;
Matches 68; Conservative 50; Mismatches 125; Indels 28; Gaps 11;

Qy 1 MYSIIMSVYNEPLNYVRDSVESILNQTLDPEFIIVID-NPSRGDLKQFLTEYSVVDNR 59
Db 1 MKISIVIVYTKR-IPALCELLESISRQTLMPYEIIIVNDAGESVVPVKALYPELP----- 54
Qy 60 IKILLNENIGLASSLNKAVISKGEYIFRMDADDISYPSRFDKQIRWEENSDFSATL 119
Db 55 IAVINLEKNSGHVAAARNAGVKEASGDCIMLCDDDDFFTPPGHIEKMAKETIADTFVHSDAE 114
Qy 120 IELDOKGNLYVKQRESNKIY-LTNDIRKMLNRSILAHPTWCVKKKVFDKLMGY--RD 176
Db 115 IVSPEEKNGTRYP--VSRKLFAYTADYEDMRVFSYV--PGSMYRRRLHDEIGYFDADV 170
Qy 177 VPVEDYDFAIRGALADFKIGLLNKVLQYRLNENGISOTNKFQYIYSAILQDFYK 232
Db 171 HNYWDWDFYLRAA-KDYRVKRVPCASVIYAFSDAGDQNSADLGAQRKOYL-----DRLS 223
Qy 233 EKSVIDITKIINY----FOEYVIKKRYTQOEL 260
Db 224 EKHLGELGELPTKNFAVLLEPEMKRREKSEM 254

RESULT 13
Y586_ANASP
ID Y586_ANASP STANDARD; PRT; 322 AA.
AC P22639;
```

```
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Putative glycosyl transferase alr2836 (EC 2.-.-.-).
GN ALR2836
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriiguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120."
RL DNA Res. 8:205-213(2001).
RN [2]
RP SEQUENCE OF 1-131 FROM N.A.
RX MEDLINE=90264305; PubMed=2111805;
RA Holland D., Wolk C.P.;
RT "Identification and characterization of heta, a gene that acts early
RT in the process of morphological differentiation of heterocysts."
RL J. Bacteriol. 172:3131-3137(1990).
CC - SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF003591; BAB74535.1; -
DR EMBL; AF031959; AAC32401.1; ALT_INIT.
DR PIR; B35391; B35391.
DR InterPro; IPRO01173; Glycos.transf_2.
DR Pfam; PF00535; Glycos.transf_2; 1.
KW Hypothetical protein; Transferase; Glycosyltransferase;
KW Complete proteome.
SQ SEQUENCE 322 AA; 36388 MW; 753C2FB59327D968 CRC64;

Query Match          9.1%; Score 144.5; DB 1; Length 322;
Best Local Similarity 20.8%; Pred. No. 0.0056;
Matches 61; Conservative 65; Mismatches 114; Indels 53; Gaps 14;

QY 1 MNYSIIMSYNEPLNYVRDSVESILNQTLDFFETIIVDNPGRGDLKQFLTE-YSYVDNR 59
DB 1 MKISVIISNYN-ARYLSRAINSVLAOTSDIE-IVIVDDGSTDSRDVITQLQEAPDK 58
QY 60 IKLLNEENIGLASSLNKAVKISKGEYIFRMDADDISYSPRFDKQIRFMEENSLDPSATL 119
DB 59 IKPIF-QANOGOGGAFNAGAAATGEVAFLOADDVKPKLQRIVEVFQTSDDVGVMMH 117
QY 120 IELIDQGNLVYKQRESNKIYLTNDIRKMLNRSILAHPTWCVKVKVFDKLMGYRDL 176
DB 118 LDIIDGNDXTI-DOASTQGPKLSIEDLASVILQTG-----NANCFPTSGLAYREVLEKVF 172
QY 172 GYRDLVPVED-----YDFAIRGALADFKIGLLNKVLLQYRLN--ENGISQTNK 217
DB 173 -----PIDPVKRWIRIADGCIYCTAFLG-----KIKTLOENLAYVRIHGANHMSAA--- 219
QY 218 FKQYIYSAILQDFYKEKSVIDITKITYQEVYIKKRY-TQOELSKYFELKST 269
DB 220 -----SATSEQAKSQAGIEMT--NQYINDFLVRIGYGARVDLSRNLYRRT 264

RESULT 14
YKCC_BACSU
```

```
ID YKCC_BACSU STANDARD; PRT; 323 AA.
AC O34319;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative glycosyl transferase ykcc (EC 2.-.-.-).
GN YKCC.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Devine K.M.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
CC - SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2. GTRB
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ002571; CAA05569.1; -
DR EMBL; Z99110; CAB13146.1; -
DR Subtilist; BG13230; ykcc.
DR InterPro; IPRO01173; Glycos.transf_2.
DR Pfam; PF00535; Glycos.transf_2; 1.
KW Hypothetical protein; Transferase; Glycosyltransferase; Transmembrane;
KW Complete proteome.
FT TRANSMEM 232 252 POTENTIAL.
FT TRANSMEM 267 287 POTENTIAL.
SQ SEQUENCE 323 AA; 37373 MW; 62078C0B6BBC3075 CRC64;

Query Match          8.8%; Score 140; DB 1; Length 323;
Best Local Similarity 22.6%; Pred. No. 0.011;
Matches 77; Conservative 57; Mismatches 116; Indels 90; Gaps 15;

QY 1 MNYSIIMSYNEPL-----NYVRDSVESILNQTLDFFETIIVDNPGRGDLKQFLTEYSYV 56
DB 5 IQYSIVVPVYNEELVTHETQYR--LKEVMDQTKENYELLFVNDG-SKDRSIELREHSLI 61
QY 57 DNRIKILLNEENIGLASSLNKAVKISKGEYIFRMDADDISYSPRFDKQIRFMEENSLDPS 116
DB 62 DPRVKLIIDFSRNGHQIATAGMDYAOGNAIVVIDADLODPPELILEMIEKWKEG---YE 118
QY 117 ATLIELIDQGNLVYKQRESNKIYLTNDIRKMLNRSILAHPTWCVKVKVFDKLMGYRDL 176
DB 119 VYVAVETKRGGETFFKKQTAAMFYRL-----LSGWTDI 151
QY 177 -VPVEDYDFAIRGALADFKIGLLNKVLLQYRLNENGISQTNKFKQYIYSAI-----LQDF 230
DB 152 DIPIDTGD- ----RLMDRKVCDEM-----RLKEK-----NPFVRLGVSWVSGSKQTAVEY 197
QY 231 YKEKSYDITKITYQEVYIKKRYTQOELS-----KYFELKSTPSITIRK---LY 278
DB 198 VDERLAGETK-----YPLKK---MLKLSMDGITTFSHKPLKLYASAGILMSGTGFLY 247
QY 279 ICILYFK-----SPLVRRLLINDINILVKLF 307
DB 248 MFIVLYKLFTDSTITGWSLLVIQLLFSGIVLLILGVIG 287

RESULT 15
AMSB_ERWAM
ID AMSB_ERWAM STANDARD; PRT; 301 AA.
AC Q46632;
DT 01-NOV-1997 (Rel. 35, Created)
```

Search completed: July 31, 2002, 19:00:10
Job time: 225 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 31, 2002, 18:51:04 ; Search time 17.53 seconds
(without alignments)
1715.685 Million cell updates/sec

Title: US-09-900-038A-1
Perfect score: 1590
Sequence: 1 MNYSIMSYNEPLNVRDS.....LINDINILVLKFGGKQSD 313

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	431.5	27.1	290	2 H64431	glycosyl transfera
2	367	23.1	333	2 H97167	glycosyltransferas
3	348.5	21.9	336	2 A97168	glycosyltransferas
4	317	19.9	333	2 B97168	glycosyltransferas
5	300.5	18.9	271	2 B84114	exopolysaccharide
6	282	17.7	340	2 T44330	glycosyl transfera
7	271	17.0	367	2 G95948	probable glycosylt
8	264	16.6	323	1 H64130	glycosyl transfera
9	251	15.8	321	2 AG2188	hypothetical prote
10	246.5	15.5	257	2 E84107	telchuronic acid b
11	221.5	13.9	329	2 AB2092	glycosyltransferas
12	221	13.9	278	2 D70036	exopolysaccharide
13	220	13.8	252	2 A69728	biosynthesis of te
14	220	13.8	604	2 E97757	hypothetical prote
15	218.5	13.7	333	2 AH2026	hypothetical prote
16	218	13.7	260	2 E90984	probable glycosyl
17	218	13.7	260	2 H85829	glycosyl transfera
18	217.5	13.7	280	2 E71703	glycosyl transfera
19	216.5	13.6	281	2 G97777	glycosyl transfera
20	215.5	13.6	318	1 E71690	minor telchoic aci
21	215.5	13.6	346	2 H81970	lacto-N-neotetraos
22	212.5	13.4	343	2 AI2091	glycosyltransferas
23	210.5	13.2	318	2 AG2189	hypothetical prote
24	208	13.1	294	2 E83022	probable glycosyl
25	206.5	13.0	267	2 C64175	hypothetical prote
26	206.5	13.0	294	2 G71148	probable glycosyl
27	205.5	12.9	334	1 G71153	hypothetical prote
28	205	12.9	349	2 D81027	lacto-N-neotetraos
29	204.5	12.9	291	2 F95015	glycosyl transfera

30	203	12.8	333	1 S70813	glycosyl transfera
31	202.5	12.7	317	2 H97888	glycosyl transfera
32	202	12.7	268	2 B97189	glycosyltransferas
33	201	12.6	344	2 A70037	capsular polysacch
34	198	12.5	298	2 B75096	glycosyl transfera
35	197	12.4	344	2 G70036	spore coat polysac
36	196.5	12.4	294	2 D64175	glycosyltransferas
37	192	12.1	900	2 E89631	galactosamine-cont
38	191.5	12.0	258	2 S61898	glycosyltransferase
39	190.5	12.0	318	2 AH2189	hypothetical prote
40	190.5	12.0	465	2 T00090	glycosyltransferas
41	189	11.9	250	2 A64099	glycosyl transfera
42	188	11.8	318	2 T50039	beta-1,4-galactosy
43	187.5	11.8	596	2 AD1876	hypothetical prote
44	186	11.7	274	2 E81319	probable glucosylt
45	185.5	11.7	996	2 AG2366	hypothetical prote

ALIGNMENTS

RESULT 1

H64431

glycosyl transferase (EC 2.4.-.-) - Methanococcus jannaschii

C:Species: Methanococcus jannaschii

C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 10-Oct-1997

C:Accession: H64431

R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blak
rson, J.D.; Sadow, P.W.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek,
Science 273, 1058-1073, 1996

A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese

A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannasc
A:Reference number: A64300; MUID:96337999

A:Accession: H64431

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-290 <BUIL>

A:Cross-references: GB:U67549; GB:L77117; NID:g1591709; PID:g1592318; TIGR:MJ1057; PI

C:Genetics:

A:Map position: FOR996513-997385

C:Keywords: glycosyltransferase

Query Match 27.1%; Score 431.5; DB 2: Length 290;
Best Local Similarity 34.9%; Pred No. 9.7e-21;
Matches 98; Conservative 58; Mismatches 102; Indels 23; Gaps 8;

Qy 4 SIIMSVYNEPLNVRDSVESILNQTLDPEFTIVIDNPSRGDLKQFLTEYSYVDNRKIL 63

Db 11 SVMATYNEPKYKESIESIXNQTKDQFXFIIXLDNPNKKAETIKNEYQQDKRIIFI 70

Qy 64 LNEENIGLASSLNKAVKISKGEYIFRMDADDISYSPRDKQIRFMEEN-SLDFSATLTIEL 122

Db 71 KNERNLGRGASRNKAVXIARGKYIALLDADDIALPKRLEKQFKYMENNRDIDLLFSWYF 130

Qy 123 IDOKNLVYKORESKYILNTDIRKMLNRSILAHPTWCVKKKVFDKMGYRDLYPVDEY 182

Db 131 IDENGNIL-KEFKPKK-YKFEIKKYFFKEHLTVAPSMVMSKILKKLYDKELINSQDY 188

Qy 183 DFAIRGALADEKIGLNLKVLQYRL--NENGISQTNKFKQYIYSAILODFYKESY---- 236

Db 189 DFWIRCIANDYKFDIEEFLLKRYIPNRDNYLSRIKKQKLSYYT-LKTHWNKKHFCNN 247

Qy 237 IDITKITYFOBYVTKKRYTQOELSXYFELKSTPSITIRKL 277

Db 248 VYFWKV--FFYSLVV-----YLFVLTPTFILKIL 275

RESULT 2

H97167

glycosyltransferase [imported] - Clostridium acetobutylicum

C:Species: Clostridium acetobutylicum

[illegible]

[illegible]

RESULT 6
T44330
glycosyl transferase homolog [imported] - *Vibrio cholerae*
C:Species: *Vibrio cholerae*
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C:Accession: T44330
R:Yamasaki, S.; Shimizu, T.; Hoshino, K.; Ho, S.T.; Shimada, T.; Takeda, Y.
Gene 237, 321-332, 1999
A:Title: the genes responsible for O-antigen synthesis of *Vibrio cholerae* O139 are close
A:Reference number: Z22749; MUID:99453293
A:Accession: T44330
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-340 <YAM>
A:CROSS-references: EMBL:AB012957; NID:q4115688; PIDN:BAA33634.1; PID:q3721684
A:Experimental source: strain O22
C:Genetics:
A:Note: wblc
C:Superfamily: *Neisseria meningitidis* glycosyl transferase A

Query Match	17.7%	Score	282;	DB 2:	Length	340;
Best Local Similarity	36.9%;	Pred.	No. 4.6e-11;			
Matches	82; Conservative	35; Mismatches	77; Indels	28; Gaps	10;	
Qy	4	SIMSVMNEPLNYVRDSVESILNQTLDFFEFIIIVDNPISRGDLKQFLTEYSVDNRKIIL	63			
Dd	9	SVMSVNGE-KYLAQAIESILNTFSDFEIIIVDDGSTDSL-SIIQAYMDKDRI-VL	65			
Qy	64	LNEENIGLASLNKAVIKSIGEYITFRMADDISVPSPRFDKQIRMEENSLSDFSATLIELI	123			
Dd	66	ISRNVKGLPYSLENAISVSXKANYIARMADDISUPERLETQLAYMENNP-DIGVC-----	119			
Qy	124	DQGNLVYKORES---NKIYLTDIRKMILNRSILA----HPTWCVKKKYPDKLMGYRDL	176			
Dd	120	--GTLAYLFRETPSKKNMCCHPDHDSLIIRLFSVCFIHPPVMIRKSVLDDL----	172			
Qy	177	VPVEDY-----DFAIRGALAD-FKIGLLNKVLLQYRLNENGIS	213			
Dd	173	VYNENFRNSQDYELWSRIAIAEKTRFYTIQKPLLFYRDTPDGIT	214			

RESULT 7
G95948

probable glycosyltransferase protein Smb21189 [imported] - *Sinorhizobium meliloti* (str. C)
C:Species: *Sinorhizobium meliloti*
C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C:Accession: G95948
R:C:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Herberich, R.
proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
A:Title: The complete sequence of the 1.683-kb pSymb megaplasmid from the N2-fixing endosymbiont of the legume *Lotus corniculatus*
A:Reference number: A95842; MUID:2136508; PMID:11481431
A:Accession: G95948
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-367 <KUR>
A:Cross-references: GB:AL591985; PIDN:CAC49255.1; PID:g15140741; GSPDB:GN00167
A:Experimental source: strain 1021, megaplasmid pSymb
R:Galibert, F.; Finan, T.M.; Long, S.R.; Punier, A.; Abola, P.; Ampe, F.; Barloy-Hubblin, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.; L. Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelau, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, E.
A:Title: The composite genome of the legume symbiont *Sinorhizobium meliloti*.
A:Reference number: A96039; MUID:21368234; PMID:11474104
A:Contents: annotation
C:Genetics:
A:Gene: Smb21189
A:Genome: plasmid

Query Match	17.0%	Score	271	DB 2	Length	367			
Best Local Similarity	33.8%	Pred. No.	2.6e-10						
Matches	74	Conservative	52	Mismatches	79	Indels	14	Gaps	11
Qy	4	SIIMSVYN-EPLNVVRDVS	ESILNQTLDPFEIIVINDP	SRGDLKQFLTEYSVVDNRKI	62				
Db	7	SIVLPVNAEP--YIAAAIESVLRQDYERLE-VTAIDGSDTRSDRILERYKKS	DSRVSI	63					
Qy	63	LLNEENTGLASSLNKAWKVISGEYIFRMDADDISYPSRFDKQIR-FMEENS	LDPSALIE	121					
Db	64	-ISRENGLATLNEGALAKANGELIARMADDIAIYPSRLSRQVALFSAEPRLA	LSGTGID	122					
Qy	122	LIDQGNLVYKORESNKIYLTNDIRKMLNLSILAHPTWCVKKKVF-DKLMG	-RDLVPV	179					
Db	123	ML--IGNRIIRGK-PNPIYRPGSRILISMFTIPMHSVTYVYNNRNVIP	EEMLRYPDNYVHA	179					
Qy	180	EDYDPAINGALAD-FKIGLLNKVLQYRLNENGISQTNK	217						
Db	180	EDFLDFRR--IADRFPPHMIIDEALVAYRIHSDSVTSKHK	216						

RESULT 8

H64130 glycosyl transferase homolog H1578 - Haemophilus influenzae (strain RD KW20)
C/Species: Haemophilus influenzae
C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 28-Jul-2000
C/Accession: H64130
R/Gleichmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage,
F.; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidmann,
D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.;
Science 269, 496-512, 1995
A/Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter
A/Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A/Reference number: A64000; MUID:95350630
A/Accession: H64130
A/Status: nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-523 <TIGR>
A/Cross-references: GB:U32832; GB:U42023; PIDN:q1574421; PIDN:AAC23227.1; PID:q1574422
C/Superfamily: Neisseria meningitidis glycosyl transferase A

Query Match	16.6%	Score 264;	DB 1;	Length 323;
Best Local Similarity	29.0%	Pred. No. 6.2e-10;		
Matches 93;	Conservative 62;	Mismatches 114;	Indels 52;	Caps 13;

E97757
hypoetical protein RC0461 [imported] - Rickettsia conorii (strain Malish 7)
C:Species: Rickettsia conorii
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 30-Sep-2001
C:Accession: E97757
R:Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; R
Science 293, 2093-2098, 2001
A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
A:Reference number: A97700; MUID:21442074; PMID:11557893
A:Accession: E97757
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-604 <KUR>
A:Cross-references: GB:AE006914; PIDN:AAL02999.1; PID:g15619533; GSPDB:GN00173
C:Genetics:
A:Gene: RC0461

Query Match 13.8%; Score 220; DB 2; Length 604;
Best Local Similarity 25.8%; Pred. No. 9e-07;
Matches 77; Conservative 65; Mismatches 118; Indels 38; Gaps 10;

QY 4 SIIMSVNEPLNVRDSEILNQTLDFEFIIIVDNPGRDLKQFLTEYSVVDNRKIL 63
DB 296 SVIIPVYNR-INWAEATKSVLIQTHNFE-ILIIDGSDDDISE-LTAICKKDKRIK-Y 351
QY 64 LNEENIGLASSLNKAVKISKGEYIFRMDADDISYPSRFDKQIRFMEENSLDPSATLIELI 123
DB 352 FHKKNEGPAARNLGKNAIGKXIAFLSDDLFYKDKIEIQLKFMEEENFISHTSYHKI 411
QY 124 DQGNLVYKQRESNKIYLTNDIRKMLNRSILAHPTCVYKKVFDKLMGYRDLVPVED-- 181
DB 412 NEKGKVIESHVSG--LFSGNVPQVQICTPIAMPTVGTTLTFOENLFPENIRSGEDCC 468
QY 182 --YFAIRCALA-----DFKGLNKNVLLQYRLNENGISQTNK 217
DB 469 LMWISASKNSIGGIDKELSKVRISGNTFTMDPNKYSVGLIN--ITSVYLNDAIYLSKFSP 526
QY 218 FK-QVIYSAILQDFYKESYIDITK--ITNYPQEVYKRRYQQELSKYFELKSTPSI 272
DB 527 FTINLLAAVTOQLLENKNEDYKKSNI5FFKNVYIQIRTYCFVTKILILLITISI 584

RESULT 15

AH2026
hypoetical protein all1766 [imported] - Anabaena sp. (strain PCC 7120)
C:Species: Anabaena sp.
A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002
C:Accession: AH2026
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
Nakazaki, N.; Shimpou, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AH2026
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-333 <KUR>
A:Cross-references: GB:BA000019; PIDN:BAW73465.1; PID:g17130856; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all1766

Query Match 13.7%; Score 218.5; DB 2; Length 333;
Best Local Similarity 23.9%; Pred. No. 5.3e-07;
Matches 75; Conservative 70; Mismatches 120; Indels 49; Gaps 13;

QY 4 SIIMSVNEPLNVRDSEILNQTLDFEFIIIVDNPGRDLKQFLTEYSVVDNRKIL 63
DB 12 SVIIPAYNCE-KTIKKTIDSVLDQSFTDFELIVDGSQDATLD---IVSQIEDSRIKI- 66

QY 64 LNEENIGLASSLNKAVKISKGEYIFRMDADDISYPSRFDKQIRFMEENSLDPSATLIEL 122
DB 296 SVIIPVYNR-INWAEATKSVLIQTHNFE-ILIIDGSDDDISE-LTAICKKDKRIK-Y 351
QY 123 IDQGNLVYKQRESNKIYLTNDIRKMLNRSIL---AHTWCVKKKVFDKLMGY-RDLVP 178
DB 127 IDEDNFLISGR---RVTLNGDVYKKLFINNFLENGSNPLIC--KEALIALGGFDESLSKA 181
QY 179 VEDYDFAIRGALADPKIGLLNK-----VLLQYRLNENGISQTNKFKQYIYSAILQDF 230
DB 182 AQDWDMLR-----LANKYSFVAVPYVOILYRVSSNSLSSNLVROEKACKQVLEKA 232
QY 231 YKEKS-----YIDITITNYPQEVYKRRYQQE-----LSKYFELKSTPSITIRK 276
DB 233 YQARPAIGNHILHLSTIANLYKYLACKALQKPYSRSGKGLAAVFKFIWQYFLYDSS---RMRR 289
QY 277 LYICLYLYFKSPLV 290
DB 290 INSTKLKLLFKALII 303

Search completed: July 31, 2002, 18:56:46
Job time: 342 sec

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	210.5	13.2	348	1	US-08-312-387B-3	Sequence 3, Appli	
2	210.5	13.2	348	1	US-08-312-387B-11	Sequence 11, Appl	
3	210.5	13.2	348	1	US-08-683-426-3	Sequence 3, Appli	
4	210.5	13.2	348	1	US-08-683-426-11	Sequence 11, Appl	
5	210.5	13.2	348	1	US-08-683-458-3	Sequence 3, Appli	
6	210.5	13.2	348	1	US-08-683-458-11	Sequence 11, Appl	
7	210.5	13.2	348	2	US-08-878-360-3	Sequence 3, Appli	
8	210.5	13.2	348	2	US-08-878-360-11	Sequence 11, Appl	
9	210.5	13.2	348	3	US-08-478-140B-3	Sequence 3, Appli	
10	210.5	13.2	348	3	US-08-478-140B-8	Sequence 8, Appli	
11	210.5	13.2	348	4	US-09-333-412-3	Sequence 3, Appli	
12	210.5	13.2	348	4	US-09-333-412-11	Sequence 11, Appl	
13	176.5	11.1	337	1	US-08-312-387B-5	Sequence 5, Appli	
14	176.5	11.1	337	1	US-08-312-387B-12	Sequence 12, Appl	
15	176.5	11.1	337	1	US-08-683-426-5	Sequence 5, Appli	
16	176.5	11.1	337	1	US-08-683-426-12	Sequence 12, Appl	
17	176.5	11.1	337	1	US-08-683-458-5	Sequence 5, Appli	
18	176.5	11.1	337	1	US-08-683-458-12	Sequence 12, Appl	
19	176.5	11.1	337	2	US-08-878-360-5	Sequence 5, Appli	
20	176.5	11.1	337	2	US-08-878-360-12	Sequence 12, Appl	
21	176.5	11.1	337	3	US-08-478-140B-5	Sequence 5, Appli	
22	176.5	11.1	337	4	US-09-333-412-5	Sequence 5, Appli	
23	176.5	11.1	337	4	US-09-333-412-12	Sequence 12, Appl	
24	161	10.1	324	1	US-08-597-236-10	Sequence 10, Appl	
25	161	10.1	324	1	US-08-746-682A-10	Sequence 10, Appl	
26	160	10.1	281	4	US-08-961-083-196	Sequence 196, App	
27	129.5	8.1	303	4	US-08-961-083-202	Sequence 202, App	

QY 4 SIIMSYNPLNVRDVSILNQTLDTEFFIIVIDNPSRGLKQFLTEYSVVDNRKIL 63
Db 6 SVLICAYNVE-KYFAQSLAAVYNQWRNLDILIVDDGSTDGL-AIAKDFQRDSRIKIL 63
QY 64 LNEENIGLASSLNKAV-KISK-----GEYIFRMDADDISYPSRFDKQIRFMEENS--LDF 115
Db 64 AQOQNSGLPSLNLGDELAKSGGGGGEYIARTDADDIASPGWIEKIVGEMEKDRSIIAM 123
QY 116 SATLELIDOK-GNLVYKQRENSKIYLT----NDIRKMLNRSILAHPTWCYKKVDFDKL 170
Db 124 GAWLEVLSEKDGRLARHHKHKWKPTRHEDTAAFPFGNPHNNTMIMRRSVIDGG 183
QY 171 MGY----RDLPVVEDYDF-----AIRGALADFKIGLLNKVLLOVRLNENGISQTNKFKQIYI 223
Db 184 LRYDTERDW--AEDYQFWVDVSKGLRLAYP-----EALVKYRLHANQVSSKHVSQRH-- 234
QY 224 SAILQDFYKEYSYIDITKITNYFOEYVVKRYT-----QOELSKYFEL--KSTPSITIRKL 277
Db 235 -EIAOGIOK-----TARNDFLQSMGFKTRFDSLEYRQTAKAAAYELPEKDLPEEDFERA 286
QY 278 YICLYLYFK 286
Db 287 RRFYQCFK 295

RESULT 6

US-08-683-458-11
; Sequence 11, Application US/08683458
; Patent No. 5798233
; GENERAL INFORMATION:
; APPLICANT: Gotschlich, Emil C.
; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
; TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/683,458
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/312,387
; FILING DATE: September 26, 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-095A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 348 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-683-458-11

Query Match 13.2%; Score 210.5; DB 1; Length 348;
Best Local Similarity 27.8%; Pred. No. 1.3e-11;

Matches 86; Conservative 55; Mismatches 123; Indels 45; Gaps 15;
QY 4 SIIMSYNPLNVRDVSILNQTLDTEFFIIVIDNPSRGLKQFLTEYSVVDNRKIL 63
Db 6 SVLICAYNVE-KYFAQSLAAVYNQWRNLDILIVDDGSTDGL-AIAKDFQRDSRIKIL 63
QY 64 LNEENIGLASSLNKAV-KISK-----GEYIFRMDADDISYPSRFDKQIRFMEENS--LDF 115
Db 64 AQOQNSGLPSLNLGDELAKSGGGGGEYIARTDADDIASPGWIEKIVGEMEKDRSIIAM 123
QY 116 SATLELIDOK-GNLVYKQRENSKIYLT----NDIRKMLNRSILAHPTWCYKKVDFDKL 170
Db 124 GAWLEVLSEKDGRLARHHKHKWKPTRHEDTAAFPFGNPHNNTMIMRRSVIDGG 183
QY 171 MGY----RDLPVVEDYDF-----AIRGALADFKIGLLNKVLLOVRLNENGISQTNKFKQIYI 223
Db 184 LRYDTERDW--AEDYQFWVDVSKGLRLAYP-----EALVKYRLHANQVSSKHVSQRH-- 234
QY 224 SAILQDFYKEYSYIDITKITNYFOEYVVKRYT-----QOELSKYFEL--KSTPSITIRKL 277
Db 235 -EIAOGIOK-----TARNDFLQSMGFKTRFDSLEYRQTAKAAAYELPEKDLPEEDFERA 286
QY 278 YICLYLYFK 286
Db 287 RRFYQCFK 295

RESULT 7

US-08-878-360-3
; Sequence 3, Application US/08878360
; Patent No. 5945322
; GENERAL INFORMATION:
; APPLICANT: Gotschlich, Emil C.
; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
; TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/878,360
; FILING DATE: 18-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/683,426
; FILING DATE:
; APPLICATION NUMBER: 08/312,387
; FILING DATE: September 26, 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-095B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 348 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-878-360-3

```
;
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-878-360-11

Query Match      13.2%; Score 210.5; DB 2; Length 348;
Best Local Similarity 27.8%; Pred. No. 1.3e-11;
Matches 86; Conservative 55; Mismatches 123; Indels 45; Gaps 15;

QY 4 SIIMSVNEPLNYVRDVSVESILNQTLDFFIIVIDNPSRGDLKQFLTEYSVVDNRKIL 63
   I::: I I I I I I I I I I I I I I I I I I I I I I I I I I I I
Db 6 SVLICAYNVE-KYFAQSLAAVNTWRNLDLIIVDDGSTDGTL-AIAKDFQRDSRIKIL 63
   : I I I I I I I I I I I I I I I I I I I I I I I I I I I I
QY 64 LNEENIGLASSLNKAV-KISK-----GEYIFRMDADDISYPSRFDKQIRPWEENS--LDF 115
   : I I I I I I I I I I I I I I I I I I I I I I I I I I I I
Db 64 AQONSLGILPSNLGDELAKSGGGGEYIARTDADDIASPGWIEKIVGEMEKDRSIAM 123
   : I I I I I I I I I I I I I I I I I I I I I I I I I I I I
QY 116 SATLIELIDQK-GNLVYKQRESNKIYLT-----NDIRKMLNRSILAHPTWCVKKKVDFKL 170
   : I I I I I I I I I I I I I I I I I I I I I I I I I I I I
Db 124 GAWLEVLSEEDGNRLARHHKHGKIWKKPTRHEDIAAFPPGPNHNTMTIMRRSVIDGG 183
   : I I I I I I I I I I I I I I I I I I I I I I I I I I I I
QY 171 MGY---RDLVPVEDYDF-----AIRGALADFKIGLLKVLQYRLNENGISQTNKFKQIYI 223
   : I I I I I I I I I I I I I I I I I I I I I I I I I I I I
Db 184 LRYDTERDW--AEDYQFWYDVSKLGRLAYYP-----EALVKYRLHANQVSSKHSVRQH-- 234
   : I I I I I I I I I I I I I I I I I I I I I I I I I I I I
QY 224 SAILQDFYKESYIDITKIINYFOEYVIKKRYT-----QOELSKYFEL--KSTPSITIRKL 277
   : I I I I I I I I I I I I I I I I I I I I I I I I I I I I
Db 235 -ETAQGIQK-----TARNDFLOSMGFKTRFDSLEYRQTAAAYELPEKDLPEDFERA 286
   : I I I I I I I I I I I I I I I I I I I I I I I I I I I I
QY 278 YICLYLYFK 286
   I I I I I I I I I I I I I I I I I I I I I I I I I I I I
Db 287 RRELYQCCK 295
   I I I I I I I I I I I I I I I I I I I I I I I I I I I I

RESULT 8
US-08-878-360-11
; Sequence 11, Application US/08878360
; Patent No. 5945322
; GENERAL INFORMATION:
; APPLICANT: Gotschlich, Emil C.
; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
; TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/878,360
; FILING DATE: 18-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/683,426
; FILING DATE:
; APPLICATION NUMBER: 08/312,387
; FILING DATE: September 26, 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-095B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 348 amino acids
```

```
;
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-878-360-11

Query Match      13.2%; Score 210.5; DB 2; Length 348;
Best Local Similarity 27.8%; Pred. No. 1.3e-11;
Matches 86; Conservative 55; Mismatches 123; Indels 45; Gaps 15;

QY 4 SIIMSVNEPLNYVRDVSVESILNQTLDFFIIVIDNPSRGDLKQFLTEYSVVDNRKIL 63
   I::: I I I I I I I I I I I I I I I I I I I I I I I I I I I I
Db 6 SVLICAYNVE-KYFAQSLAAVNTWRNLDLIIVDDGSTDGTL-AIAKDFQRDSRIKIL 63
   : I I I I I I I I I I I I I I I I I I I I I I I I I I I I
QY 64 LNEENIGLASSLNKAV-KISK-----GEYIFRMDADDISYPSRFDKQIRPWEENS--LDF 115
   : I I I I I I I I I I I I I I I I I I I I I I I I I I I I
Db 64 AQONSLGILPSNLGDELAKSGGGGEYIARTDADDIASPGWIEKIVGEMEKDRSIAM 123
   : I I I I I I I I I I I I I I I I I I I I I I I I I I I I
QY 116 SATLIELIDQK-GNLVYKQRESNKIYLT-----NDIRKMLNRSILAHPTWCVKKKVDFKL 170
   : I I I I I I I I I I I I I I I I I I I I I I I I I I I I
Db 124 GAWLEVLSEEDGNRLARHHKHGKIWKKPTRHEDIAAFPPGPNHNTMTIMRRSVIDGG 183
   : I I I I I I I I I I I I I I I I I I I I I I I I I I I I
QY 171 MGY---RDLVPVEDYDF-----AIRGALADFKIGLLKVLQYRLNENGISQTNKFKQIYI 223
   : I I I I I I I I I I I I I I I I I I I I I I I I I I I I
Db 184 LRYDTERDW--AEDYQFWYDVSKLGRLAYYP-----EALVKYRLHANQVSSKHSVRQH-- 234
   : I I I I I I I I I I I I I I I I I I I I I I I I I I I I
QY 224 SAILQDFYKESYIDITKIINYFOEYVIKKRYT-----QOELSKYFEL--KSTPSITIRKL 277
   : I I I I I I I I I I I I I I I I I I I I I I I I I I I I
Db 235 -ETAQGIQK-----TARNDFLOSMGFKTRFDSLEYRQTAAAYELPEKDLPEDFERA 286
   : I I I I I I I I I I I I I I I I I I I I I I I I I I I I
QY 278 YICLYLYFK 286
   I I I I I I I I I I I I I I I I I I I I I I I I I I I I
Db 287 RRELYQCCK 295
   I I I I I I I I I I I I I I I I I I I I I I I I I I I I

RESULT 9
US-08-478-140B-3
; Sequence 3, Application US/08478140B
; Patent No. 6127153
; GENERAL INFORMATION:
; APPLICANT: JOHNSON, KARL F.
; APPLICANT: BUCZALA, STEPHANIE L.
; TITLE OF INVENTION: METHOD OF TRANSFERRING AT LEAST TWO
; TITLE OF INVENTION: SACCHARIDE UNITS WITH A POLYGLYCOSYLTRANSFERASE, A
; TITLE OF INVENTION: POLYGLYCOSYLTRANSFERASE AND GENE ENCODING A
; TITLE OF INVENTION: POLYGLYCOSYLTRANSFERASE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,140B
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Laura A. Coruzzi
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7188-017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 3:
```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 348 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-478-140B-3

Query Match 13.2%; Score 210.5; DB 3; Length 348;
Best Local Similarity 27.8%; Pred. No. 1.3e-11;
Matches 86; Conservative 55; Mismatches 123; Indels 45; Gaps 15;

QY 4 STIMSYNEPLNYVRDSEVILNQTLTDFEFLIVDNPGRDLKQFLTEYSVVDNRKIL 63
Db 6 SVLICAYNVE-KYFAQSLAAVYVQWNRNLDILIVDDGSDGTL-AIAKDFQRDSRIKIL 63
QY 64 LNEENIGLASSLNKAV-KISK-----GEYIFRMDADDISYPSRFDKQIRFMEENS--LDF 115
Db 64 AQOQNSGLTSPINIGLDELAKSGGGGEYIARTDADDIASPGWIEKIVGEMEKDRSIAM 123
QY 116 SATLIELIDQK-GNLVYKQRESNKIYLT----NDIRKMLNRSILAHPTWCVKKVFQKL 170
Db 124 GAWLEVLSEKDGRLARHHKHKWKPTRHEDIAAEPFGNPIHNTMTIMRRSVIDGG 183
QY 171 MGY----RLVPVEDYDF-----AIRGALADFKIGLLNKVLLQYRLNENGISQTNKFKQYIY 223
Db 184 LRYDTERDW--AEDYQFWDVSKLGLRLAYP-----EALVKYRLHANQVSSKHSHVRQH-- 234
QY 224 SAILQDFYKEKSYIDITKITNYFOEYVVKRYT---OOELSKYFEL--KSTPSITIRKL 277
Db 235 -EIAQGIQK-----TARNDFLOSNGFKTRFDSLEYRQTAKAAAYELPEKDLPEEDFERA 286
QY 278 YICLYLYFK 286
Db 287 RRFLYQCFK 295

RESULT 10
US-08-478-140B-8
; Sequence 8, Application US/08478140B
; Patent No. 6327153
; GENERAL INFORMATION:
; APPLICANT: JOHNSON, KARL F.
; APPLICANT: ROTH, STEPHEN
; APPLICANT: BUCZALA, STEPHANIE L.
; TITLE OF INVENTION: METHOD OF TRANSFERRING AT LEAST TWO
; TITLE OF INVENTION: SACCHARIDE UNITS WITH A POLYGLYCOSYLTRANSFERASE, A
; TITLE OF INVENTION: POLYGLYCOSYLTRANSFERASE AND GENE ENCODING A
; TITLE OF INVENTION: POLYGLYCOSYLTRANSFERASE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,140B
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Laura A. Coruzzi
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7188-017
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864

; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 348 amino acids
; TYPE: amino acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ANTI-SENSE: NO
US-08-478-140B-8

Query Match 13.2%; Score 210.5; DB 3; Length 348;
Best Local Similarity 27.8%; Pred. No. 1.3e-11;
Matches 86; Conservative 55; Mismatches 123; Indels 45; Gaps 15;

QY 4 STIMSYNEPLNYVRDSEVILNQTLTDFEFLIVDNPGRDLKQFLTEYSVVDNRKIL 63
Db 6 SVLICAYNVE-KYFAQSLAAVYVQWNRNLDILIVDDGSDGTL-AIAKDFQRDSRIKIL 63
QY 64 LNEENIGLASSLNKAV-KISK-----GEYIFRMDADDISYPSRFDKQIRFMEENS--LDF 115
Db 64 AQOQNSGLTSPINIGLDELAKSGGGGEYIARTDADDIASPGWIEKIVGEMEKDRSIAM 123
QY 116 SATLIELIDQK-GNLVYKQRESNKIYLT----NDIRKMLNRSILAHPTWCVKKVFQKL 170
Db 124 GAWLEVLSEKDGRLARHHKHKWKPTRHEDIAAEPFGNPIHNTMTIMRRSVIDGG 183
QY 171 MGY----RLVPVEDYDF-----AIRGALADFKIGLLNKVLLQYRLNENGISQTNKFKQYIY 223
Db 184 LRYDTERDW--AEDYQFWDVSKLGLRLAYP-----EALVKYRLHANQVSSKHSHVRQH-- 234
QY 224 SAILQDFYKEKSYIDITKITNYFOEYVVKRYT---OOELSKYFEL--KSTPSITIRKL 277
Db 235 -EIAQGIQK-----TARNDFLOSNGFKTRFDSLEYRQTAKAAAYELPEKDLPEEDFERA 286
QY 278 YICLYLYFK 286
Db 287 RRFLYQCFK 295

RESULT 11
US-09-333-412-3
; Sequence 3, Application US/09333412
; Patent No. 6342382
; GENERAL INFORMATION:
; APPLICANT: Gotschlich, Emil C.
; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
; OLIGOSACCHARIDES, AND GENES ENCODING THEM
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/333,412
; FILING DATE: 15-Jun-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/312,387
; FILING DATE: July 7, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-095


```
, , REGISTRATION NUMBER: 26,742
, , REFERENCE/DOCKET NUMBER: 600-1-095
, , TELECOMMUNICATION INFORMATION:
, , TELEPHONE: 201 487-5800
, , TELEFAX: 201 343-1684
, , TELEX: 133521
, , INFORMATION FOR SEQ ID NO: 11:
, , SEQUENCE CHARACTERISTICS:
, , LENGTH: 348 amino acids
, , TYPE: amino acid
, , TOPOLOGY: linear
, , MOLECULE TYPE: protein
, , SEQUENCE DESCRIPTION: SEQ ID NO: 11:
US-09-333-412-11
```

```

US-09-533-412-11

Query Match      13.2%; Score 210.5; DB 4; Length 348;
Best Local Similarity 27.8%; Pred. No. 1.3e-11;
Matches 86; Conservative 55; Mismatches 123; Indels 45; Gaps 15;

QY   4  SILMSVYNPLNVRDVSISILNGLTLDFERIIVIDNPGRDLQKLFTEYSVVDNRKIL 63
      |:::| | | |:::| | | |:::| | | |:::| | | |:::| | | |:::|
Db   6  SVLICAVNVE-KYPAQSALAAVYVWTRNLDILIVDDGSDGTLL-AIAKQFOKRSRIKL 63
      |:::| | | |:::| | | |:::| | | |:::| | | |:::| | | |:::|

QY   64  LNEENIGLASSLNKAV-KISK-----GEYIFRMDADDISYSPRFDKOIRFMEENS--LDF 115
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   64  AQONSGLIPLSLNGLDELAKSGGGGGGEYIARTDADDIASPGWIEKIVGEMEKORSIIAM 123
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY   116  SATLIELIDOK-GLNLYVKQRESNKIYLT----DIRKMLLNRSRLTAHPCTCVKKKVFDKL 170
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   124  GAWLEVLSEKDCGNRLARHHKHGKIWKPPRTHEDIAAEPFGNPIHNNTMTMRRSVIDGG 183
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY   171  MGY---RDLVPVEDYDF-----AIRGALADPKIGILNKVLQYRLNENGISQTNKFKQYIY 223
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   184  LRYDTERDW--AEDYQWYDVSKLGRLAYP-----EALVKYRLHANOVSSKHSVRQH-- 234
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY   224  SAILQDYKPKSKYIDITKINYNFOEYVKKRYT-----QOELSKYFEL--KSTPSITIRKL 277
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   235  -ETAAQGIQK-----TARNDFLQSMGFKTRFDSLEYRQTKAAAYELPEKOLPEEDPERA 286
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY   278  YICLYLYFK 286
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   287  RRFLYQCCK 295
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

287 NRFLDQCR 293

RESULT 13
US-08-312-387B-5
; Sequence 5, Application US/08312387B
; Patent No. 5545553
; GENERAL INFORMATION:
; APPLICANT: Gotschlich, Emil C.
; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
; TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/312.387B
; FILING DATE: July 7, 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26.742

```

```

; REFERENCE/DOCKET NUMBER: 600-1-095
;
; TELECOMMUNICATION INFORMATION:
;
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
;
; INFORMATION FOR SEQ ID NO: 5:
;
; SEQUENCE CHARACTERISTICS:
;
; LENGTH: 337 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-312-387B-5

```

Query Match	11.1%	Score 176.5	DB 1	Length 337
Best Local Similarity	22.6%	Pred. No. 1.6e-08		
Matches	62	Conservative 60	Mismatches 93	Indels 59
Gaps	13			
Qy	4	SIIMSVNEPLNYVRDVS	ESILNLTLDTFEPIIVIDNP	SRGDLKQFLTEYSVDNRKIL 63
Db				
6	SVLCAYNAE-KYFAQSLAAVVGOTWRNLD-ILIVDGGSTGDTPTAIAHRHFQDGRIRII 63			
Qy	64	LNEENIGLASSLN----	KAVKISGEIYFRMDADDISYPSR	EDKQIRFME-ENSLDFSAT 118
Db				
64	SNPRNLGFIASLGLDELA	KSGGGEIYARTDADDIASPGWIEKIVGEMEKDRSIIANGA 123		
Qy	119	LIELIDOKGNLVYKQRESNKIYLTNDIRKMLNLSIAHPTWCVKKK--VFEDKLMGYRDL 176		
Db	124	WLEVLSEEN-----	NKSVLA----	ATARNCAIWDKPTRHEDI 156
Qy	177	VPVEDYDFAIRGALADFKIGLLNKVLQYRLNENGISQTNKF-KOYIYSAILOQFYEKS 235		
Db	157	VAVFPFGNPITH-----	NTMIWRSVIDGGL--	REDPAYIHA---EDY---KF 196
Qy	236	YIDITKTKTN--YFOEYVIKKRYTQOELS	KYFELK 267	
Db	197	WYEAKGLRLAYYP	EALVKYRFHODOTS	SSKYNLO 230

```

RESULT 14
US-08-312-387B-12
; Sequence 12, Application US/08312387B
; Patent No. 5545553
; GENERAL INFORMATION:
; APPLICANT: Gotschlich, Emil C.
; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
; TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/312.387B
; FILING DATE: July 7, 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-095
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 12:

```

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 337 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-312-387B-12

Query Match 11.1%; Score 176.5; DB 1; Length 337;
Best Local Similarity 22.6%; Pred. No. 1.6e-08;
Matches 62; Conservative 60; Mismatches 93; Indels 59; Gaps 13;

Qy 4 SLIMSVNEPLNYVRDVS ESI L N G T L T D F E I I V I D N P S R G D L A K Q F L T E Y S V D N R K I L 63
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 6 S V L I C A Y N A E - K Y F A Q S L A A V G G T W R N L D - I L I V D G S T D G T P A I R A H F Q E Q D G R I R I I 63
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |

Qy 64 L N E E N I G L A S S L N - - - - K A V K I S K G E Y I F R D M A D D I S Y P S R F D K Q I R F M E - E N S L D F S A T 118
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 64 S N P R N L G F I A S L N I G L D E L A K S G G G E V I A R T D A D D I A S P G W I E K I V G E M E K D R S I I A M G A 123
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |

Qy 119 L I E L I D Q K N L V Y K Q R E S N K I Y L T N D I R K M L L N R S I L A H T P W C V K K - - V P D K L M G Y R D L 176
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 124 W L E V L S E E N - - - - - - - - - - N K S V L A - - - - - A I A R N G A I W D K P T R H E D I 156
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |

Qy 177 V P V E D Y D F A I R G A L D F K I G L N K V L L Q Y R L N E N G I S Q T N K F - K O Y I Y S A I L Q D F Y K E S 235
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 157 V A V F P F G N P I H N - - - - - - - - - - N T M I M R S V I D G L - - - - R F D P A Y I H A - - - E D Y - - - K F 196
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |

Qy 236 Y I D T K T I N - - Y F O E Y V I K R Y T Q O E L S K Y F E L K 267
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 197 W Y E A G I G R L A Y Y P E A L V K Y R F H Q D Q T S S K Y N L Q 230
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |

```

```

RESULT 15
US-08-683-426-5
: Sequence 5, Application US/08683426
: Patent No. 5705367
: GENERAL INFORMATION:
: APPLICANT: Gotschlich, Emil C.
: TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
: TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM
: NUMBER OF SEQUENCES: 12
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Klauber & Jackson
: STREET: 411 Hackensack Avenue
: CITY: Hackensack
: STATE: New Jersey
: COUNTRY: USA
: ZIP: 07601
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/683.426
: FILING DATE:
: CLASSIFICATION: 536
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/312,387
: FILING DATE: September 26, 1994
: CLASSIFICATION: 536
: ATTORNEY/AGENT INFORMATION:
: NAME: Jackson Esq., David A.
: REGISTRATION NUMBER: 26,742
: REFERENCE/DOCKET NUMBER: 600-1-095B
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 201 487-5800
: TELEFAX: 201 343-1684
: TELEX: 133521
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 337 amino acids

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 31, 2002, 18:50:19 ; Search time 30.44 seconds
(without alignments)
1142.119 Million cell updates/sec

Title: US-09-900-038a-1

Perfect score: 1590

Sequence: 1 MNYSIIMSVYNEPLNVRDS.....LINDINILVLKFGGEKQSD 313

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_032802.*

- 1: /SIDSI/gcgdata/hold-geneq/geneq-emb1/AA1980.DAT.*
- 2: /SIDSI/gcgdata/hold-geneq/geneq-emb1/AA1981.DAT.*
- 3: /SIDSI/gcgdata/hold-geneq/geneq-emb1/AA1982.DAT.*
- 4: /SIDSI/gcgdata/hold-geneq/geneq-emb1/AA1983.DAT.*
- 5: /SIDSI/gcgdata/hold-geneq/geneq-emb1/AA1984.DAT.*
- 6: /SIDSI/gcgdata/hold-geneq/geneq-emb1/AA1985.DAT.*
- 7: /SIDSI/gcgdata/hold-geneq/geneq-emb1/AA1986.DAT.*
- 8: /SIDSI/gcgdata/hold-geneq/geneq-emb1/AA1987.DAT.*
- 9: /SIDSI/gcgdata/hold-geneq/geneq-emb1/AA1988.DAT.*
- 10: /SIDSI/gcgdata/hold-geneq/geneq-emb1/AA1989.DAT.*
- 11: /SIDSI/gcgdata/hold-geneq/geneq-emb1/AA1990.DAT.*
- 12: /SIDSI/gcgdata/hold-geneq/geneq-emb1/AA1991.DAT.*
- 13: /SIDSI/gcgdata/hold-geneq/geneq-emb1/AA1992.DAT.*
- 14: /SIDSI/gcgdata/hold-geneq/geneq-emb1/AA1993.DAT.*
- 15: /SIDSI/gcgdata/hold-geneq/geneq-emb1/AA1994.DAT.*
- 16: /SIDSI/gcgdata/hold-geneq/geneq-emb1/AA1995.DAT.*
- 17: /SIDSI/gcgdata/hold-geneq/geneq-emb1/AA1996.DAT.*
- 18: /SIDSI/gcgdata/hold-geneq/geneq-emb1/AA1997.DAT.*
- 19: /SIDSI/gcgdata/hold-geneq/geneq-emb1/AA1998.DAT.*
- 20: /SIDSI/gcgdata/hold-geneq/geneq-emb1/AA1999.DAT.*
- 21: /SIDSI/gcgdata/hold-geneq/geneq-emb1/AA2000.DAT.*
- 22: /SIDSI/gcgdata/hold-geneq/geneq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	272	17.1	268	21	Enzyme EPSH involv
2	272	17.1	268	21	Amino acid sequenc
3	240.5	15.1	274	22	C glutamicum prote
4	234	14.7	301	21	Campylobacter jeju
5	223	14.0	209	21	Amino acid sequenc
6	216	13.6	706	22	Enterococcus faeca
7	216	13.6	715	22	Enterococcus faeca
8	214	13.5	269	21	Cps9G protein whic
9	213	13.4	260	20	Sugar transferase
10	210.5	13.2	348	17	N. gonorrhoeae gly
11	210.5	13.2	348	18	Neisseria polyglyc

12	210	13.2	322	21	AA1568974	CpsII protein whic
13	200	12.6	270	21	AA1972033	Campylobacter jeju
14	198	12.5	298	22	AA1963113	Putative glycosylt
15	193.5	12.2	327	21	AA154095	Enzyme EPSJ involv
16	193.5	12.2	327	21	AA143797	Amino acid sequenc
17	187	11.8	332	21	AA168962	Cps2J protein whic
18	186.5	11.7	345	22	AA190140	C glutamicum prote
19	182.5	11.5	329	22	AA147427	ESPM. Lactococcus
20	180.5	11.4	302	21	AA197213	Campylobacter jeju
21	176.5	11.1	337	17	AA191314	N. gonorrhoeae gly
22	176.5	11.1	337	18	AA106579	Lipo-oligosacchari
23	168	10.6	277	19	AA180728	S. pneumoniae gly
24	168	10.6	702	21	AA196212	P. multocida hyalu
25	165.5	10.4	972	20	AA106212	Pasteurella multoc
26	165.5	10.4	972	21	AA143099	P. multocida hyalu
27	165	10.4	334	21	AA168963	Cps2K protein whic
28	165	10.4	336	21	AA154098	Enzyme EPSM involv
29	165	10.4	336	21	AA143800	Amino acid sequenc
30	165	10.4	965	21	AA196213	P. multocida chond
31	164	10.3	278	21	AA168976	Cps1K protein whic
32	161	10.1	324	18	AA14078	S. thermophilus exo
33	161	10.1	324	18	AA122177	S. thermophilus exo
34	161	10.1	328	21	AA181720	Streptococcus pneu
35	160	10.1	281	19	AA161236	Streptococcus pneu
36	159.5	10.0	332	22	AA147426	EpsN. Lactococcus
37	158.5	10.0	330	21	AA154080	Enzyme EPS6 involv
38	158.5	10.0	330	21	AA143782	Amino acid sequenc
39	157.5	9.9	322	21	AA168975	Cps1J protein whic
40	156.5	9.8	322	21	AA154071	Enzyme EPS7 whic
41	156.5	9.8	322	21	AA143773	Amino acid sequenc
42	150	9.4	973	22	AB102248	Novel human diagn
43	150	9.4	973	22	AB102248	Novel human diagn
44	150	9.4	973	22	AB102248	Novel human diagn
45	147	9.2	316	21	AA154072	Enzyme EPS8 whic

ALIGNMENTS

RESULT 1
AA154093
ID AA154093 standard; Protein; 268 AA.
XX AC AA154093;
DT 27-MAR-2000 (first entry)
XX DE Enzyme EPSH involved in exopolysaccharide biosynthesis.
XX KW Exopolysaccharide; EPS; ESP enzyme; EPSA; EPSB; EPSD; EPSF; EPSG; EPSH; EPSJ; EPSK; Lactobacillus delbrueckii bulgaricus;
KW activated D-galactose pyranose; saccharide; beta-glycosyltransferase;
KW transcription attenuator; glucosyl-phospho-transferase;
KW alpha-glycosyltransferase; glucosyltransferase; EPSL; EPSM;
KW alpha-glycosyltransferase; EPS polymerase; glycosyltransferase; EPSN;
KW phosphofuranose; transporter; food; fermented milk product; yoghurt;
KW cheese; flavour stability; organoleptic property.
XX OS Lactobacillus delbrueckii bulgaricus.
XX PN WO9962316-A2.
XX PD 09-DEC-1999.
XX PF 22-APR-1999; 99WO-EP02841.
XX PR 22-APR-1998; 98EP-0201310.
XX PR 22-APR-1998; 98EP-0201311.
XX PR 22-APR-1998; 98EP-0201312.
XX PA (NEST) SOC PROD NESTLE SA.
XX PI Stingele F, Germond JE, Lamothe G;


```
XX New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
XX Example 3: Seq ID No 4950; 51lpp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 706 AA;
XX
Query Match 13.6%; Score 216; DB 22; Length 706;
Best Local Similarity 23.3%; Pred. No. 3.3e-11;
Matches 79; Conservative 64; Mismatches 138; Indels 58; Gaps 10;
QY 4 SIIMSVYNEPLNVYRDSVESILNQTLDFFFIIVIDNPSRGDLKQFLTEYSVVDNRKIL 63
Db 183 siampvynveekwlrldcidsilnqvynwclmaddastdpnvkklitvqqdlerirv 242
QY 64 LNEENIGLASSLNKAVKISRGEYIFRMDADDISVPSRFDKQIRFMEEN-SLDFSATLIEL 122
Db 243 freqnghiseatsnalaatgefvalldnddelainafeyvkvlnenpeldiysdedk 302
QY 123 IDKGNLVYKQRESNKIYLTNDIRKMLNRSILAHPTWCVKKKVDFKLMGYR-DLVVPVED 181
Db 303 idmdgn-----rsdpafkpdwspdlilgtynyish-lgvyrirsileiegfrkgyegsqd 355
QY 182 YDFAIRGALADFK--IGLLNKVLQYRLNENGISQTNKFKQYIYSA----- 225
Db 356 ydlvlftrfektktkerithipkvlywrmplststavgdsgskgyafeaglravqdalvrri 415
QY 226 -----ILQDFYKEK-----SYIDITKITNYFQEVYVKKRYTQOELS 261
Db 416 ngathgaanglydvvydiesekivsiilptkngykdqrcvssile---ktygnyei- 471
QY 262 KYPELAKSTPSITRKLYICLYLFPKSPVLRRLINDINI 300
Db 472 ----imadngstqpkmhe-lyakfeqlpgrffvesidi 505
RESULT 7
AAU34918
ID AAU34918 standard; Protein: 715 AA.
XX
XX AAU34918;
XX
DT 14-FEB-2002 (first entry)
XX
DE Enterococcus faecalis cellular proliferation protein #205.
XX Antisense; prokaryotic cellular proliferation protein;
KW antibiotic; antibacterial; drug design.
XX
OS Enterococcus faecalis.
XX
```

```
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US09180.
XX
XX 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 23-OCT-2000; 2000US-242578P.
PR 22-DEC-2000; 2000US-253625P.
PR 16-FEB-2001; 2001US-269308P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
XX WPI; 2001-611495/70.
XX N-PSDB; AAS52777.
XX
XX New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
XX Example 3: Seq ID No 10511; 51lpp; English.
XX
XX The invention relates to antisense inhibitors of genes essential to
XX prokaryotic cellular proliferation, their use in identifying the
XX genes, their use in the discovery of novel antibiotics, the essential
XX genes themselves and the encoded proteins. The prokaryotes used are
XX Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
XX pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
XX invention is also useful for the identification of potential new targets
XX for antibiotic development. The antisense nucleic acids can also be used
XX to identify proteins used in proliferation, to express these proteins,
XX and to obtain antibodies capable of binding to the expressed proteins.
XX The proteins can be used to screen compounds in rational drug discovery
XX programmes. The antisense nucleic acid sequence is also useful to screen
XX for homologous nucleic acids which are required for cell proliferation in
XX a wide variety of organisms. The present sequence represents an
XX essential prokaryotic cellular proliferation protein.
XX Note: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic
XX format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 715 AA;
XX
Query Match 13.6%; Score 216; DB 22; Length 715;
Best Local Similarity 23.3%; Pred. No. 3.3e-11;
Matches 79; Conservative 64; Mismatches 138; Indels 58; Gaps 10;
QY 4 SIIMSVYNEPLNVYRDSVESILNQTLDFFFIIVIDNPSRGDLKQFLTEYSVVDNRKIL 63
Db 185 siampvynveekwlrldcidsilnqvynwclmaddastdpnvkklitvqqdlerirv 244
QY 64 LNEENIGLASSLNKAVKISRGEYIFRMDADDISVPSRFDKQIRFMEEN-SLDFSATLIEL 122
Db 245 freqnghiseatsnalaatgefvalldnddelainafeyvkvlnenpeldiysdedk 304
QY 123 IDKGNLVYKQRESNKIYLTNDIRKMLNRSILAHPTWCVKKKVDFKLMGYR-DLVVPVED 181
Db 305 idmdgn-----rsdpafkpdwspdlilgtynyish-lgvyrirsileiegfrkgyegsqd 357
QY 182 YDFAIRGALADFK--IGLLNKVLQYRLNENGISQTNKFKQYIYSA----- 225
Db 358 ydlvlftrfektktkerithipkvlywrmplststavgdsgskgyafeaglravqdalvrri 417
QY 226 -----ILQDFYKEK-----SYIDITKITNYFQEVYVKKRYTQOELS 261
Db 418 ngathgaanglydvvydiesekivsiilptkngykdqrcvssile---ktygnyei- 473
```


Qy	4	SIMSVTNEPLNVRDVS	ESILNQTLTDFEIIIVINPSRGDLKQFLTEYSVVVDNRKIL	63
			: :: : : : : :	
Db	7	silmpvnyngaktii-svesi	ihqsyqdf-vlyliddcstdtfsinsryknnqkiril	64
Qy	64	LNENIGLASSLNKAVKIS	GEYIFRMADDDISYPSRFDKQIRFMEENSIDFSATLIETL	123
		: : : :: : : : : :	: : : : : : : : :	
Db	65	rnktnlgaesrnygiemat	gkyisfcdaddlhcklerqievlnneevdvcsnyyvl	124
Qy	124	DQKGNLYVKQESNKIYLT	NDIRKMLNRSILAHPTWCVKK-----KVFDKLMGYRDLVPV	179
		: : : : : : : :	: : : : : : : :	
Db	125	dnorniv--gevnaphvin	-yrkmlm-knyignlgtynanlkgfygkkgihedylmw	179
Qy	180	EDYDPAIRGALADFKTGL	LNNKVLLOYRLNENGISQTNKPK--QYIYSAILQDFYKESVI	237
		: : : : : : : : :	: : : : : : : :	
Db	180	lelinktngaic-----	lqdnlayymrsnmsis-gnkikaakwtws-----iyrehhl	227
Qy	238	DITKITNYFOEYV--IK	KRYTQOELSK 262	
		: : : : : : :	: : : : : : :	
Db	228	sfpktlyfilyasngymk	ithsllrr 255	
RESULT	10			
AAR91311				
ID	AAR91311	standard; Protein; 348 AA.		
XX				
AC	AAR91311;			
XX				
DT	09-JUL-1996	(first entry)		
XX				
DE		N. gonorrhoeae glycosyltransferase LgtA.		
XX				
KW		Glycosyltransferase; lipo-oligosaccharide; lgt gene; LOS locus;		
KW		vaccine.		
XX				
OS		Neisseria gonorrhoeae strain F62.		
XX				
PN		WO9610085-A1.		
XX				
PD		04-APR-1996.		
XX				
PF		25-SEP-1995; 95WO-US12317.		
XX				
PR		26-SEP-1994; 94US-0312387.		
XX		(UYRQ) UNIV ROCKEFELLER.		
PA				
XX		Gotschlich EC;		
PI				
XX		WPI; 1996-200924/20.		
DR		N-PSDB; AAT14061.		
XX				
PT		Nucleic acids encoding glycosyl transferase(s) - used in the		
PT		diagnosis of infection with Neisseria and for the biosynthesis of		
PT		oligo:saccharide(s)		
XX				
PS		Claim 8; Fig 2b; 81pp; English.		
XX				
CC		5 Glycosyltransferases (AAR91311-15) are products of the lgt locus		
CC		(AAT14061) of Neisseria gonorrhoeae strain F62; Glycosyltransferase		
CC		LgtA (AAR91311) can be obtd. by expression of the lgtA coding		
CC		sequence in recombinant host cells. A method for adding GalNac		
CC		or GlcNAc beta1-3 to Gal comprises contacting a reaction mixture		
CC		contg. activated GalNac or GlcNAc to an acceptor moiety comprising		
CC		a Gal residue in the presence of LgtA. Oligosaccharides can be		
CC		produced that, when attached to non-toxic lipids, are useful for		
CC		Neisseria vaccine prepn. Blood group core oligosaccharides, and		
CC		mnics of lacto-N-neotetraose, gangliosides and saccharide portions		
CC		of globoglycolipids can also be produced using the enzymes.		
XX				
SQ	Sequence	348 AA;		

Query Match	13.2%	Score 210.5	DB 17	Length 348
Best Local Similarity	27.8%	Pred. No. 3.8e-11		
Matches	86	Conservative 55	Mismatches 123	Indels 45
Gaps				
QY	4	SIIMSVNEPLNLYVRDSVESITNLTQTFDEFFIIVIDNPISRGDLKQFLTEYSVVVDNRKIL	63	
DB	6	svllcaynve-kyfaqsiaavvngtwnldilivddgstdgtl-atakdqkdrdsrikil	63	
QY	64	LNEENIGLASSNAKV-KISK-----GEYIFRMDADDISYPSRFDKQIRPMEENS--LDF	115	
DB	64	aqansglipnlgldelaksqgggeyiairtaddaddiaspgwlektivgemekdrslia	123	
QY	116	SATLIELIDOK-GNLVYKQRESNKIYIT----NDIRKMLLNRSILAHPTWCVKKKVDFDKL	170	
DB	124	gawlieviseekdgnrlarhnhkhgkiwktptrhediaaaffpgnpihntmimrsvldgg	183	
QY	171	MGY---RDLPVVEDYDF----AIRGALADPKIGLLNKVILLOYRLNENGISQTNKFKQYIY	223	
DB	184	lrydterdw--aedqfydwvsklgrlayyp-----ealvkylrhanqvsskshsvrgh--	234	
QY	224	SAILODFVKEKSYIDITKITYNFOEYVKKRYT----QQELSKYFEL--KSTPSITIRKL	277	
DB	235	-elaagqik-----farndfqlsmgfktrfidsleyrqtkaaayelpdkdlpeedfera	286	
QY	278	YICLYLYFK	286	
DB	287	rrflyqcfc	295	
RESULT	11			
AAW06576				
ID	AAW06576	standard; Protein; 348 AA.		
AC	AAW06576			
XX				
XX	21-MAR-1997	(first entry)		
DT				
XX				
DE				
XX				
XX	Neisseria polyglycosyltransferase.			
KW	Polyglycosyltransferase; N-acetylglucosaminyl transferase;			
KW	N-acetylgalactosaminyl transferase; lipo-oligosaccharide.			
XX				
OS	Neisseria gonorrhoeae ATCC 33084.			
XX				
PN	W09640971-A1.			
XX				
PD	19-DEC-1996.			
XX				
PF	03-JUN-1996; 96WO-US08323.			
XX				
PR	07-JUN-1995; 95US-0478140.			
XX				
PA	(NEOS-) NEOSE TECHNOLOGIES INC.			
XX				
PI	Buczala SL, Johnson KF, Roth S;			
XX				
DR	WPI; 1997-052351/05.			
XX	N-PSDB; AAT49230.			
XX				
PT	Transfer of at least 2 saccharide units using			
PT	poly:glycosyl:transferase - isolated from N. gonorrhoeae, catalyses			
PT	the addition of both GlcNAc and GalNAc di:saccharide(s) units to a			
XX	single galactose moiety			
XX				
PS	Disclosure; Fig 2A-C; 38pp; English.			
XX				
CC	A novel polyglycosyltransferase (PGTase) (AAW06576) from Neisseria			
CC	gonorrhoeae ATCC 33084 catalyses the stereospecific conjugation of			
CC	2 specific activated saccharide units (e.g. UDP-GlcNAc, UDP-GalNAc,			
CC	UDP-Gal) to specific acceptors having a galactose moiety at a			
CC	non-reducing end. It is the first PGTase reported to be capable of			
CC	transfer of more than one different saccharide moiety. The PGTase			
CC	is encoded by nucleotides 445-1488 of a lipo-oligosaccharide gene			

```
CC (AA749230). It can be produced in transformed host cells and used in
XX oligosaccharide prodn.
SQ Sequence 348 AA;

Query Match 13.2%; Score 210.5; DB 18; Length 348;
Best Local Similarity 27.8%; Pred. No. 3.8e-11;
Matches 86; Conservative 55; Mismatches 123; Indels 45; Gaps 15;

QY 4 SIIMSYNEPLNVVRDSVESILNQTLDFEIIIVIDNPSRGDLKQFLTEYSVVDNRKIL 63
Db 1::: || | : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
6 swicaynve-kyfaqsiaavvntqwnrldilivdgsdgtl-aiakdfqrdrikil 63

QY 64 LNEENIGLASSLNKAV-KISK-----GEYIFRMDADDISYPSRFDKQIRFMEENS--LDF 115
Db 64 aqaqngslipsiniglidelaksggggeyiaartdaddiaspwiekivgemekdrsilam 123

QY 116 SATLIELIDQK-GNLVYKQRESNKIYLT-----NDIRKMLNRSILAHPTWCVKKKVFDKL 170
Db 124 gawlevlseekdgnrlarhhkgkiwkptrhediaaaffpgnphntmimrrsvldgg 183

QY 171 MGY--RDLPVVEDYDF-----AIRGALADFKIGLLNKVLQYLRNENGISQFNKFKQYIY 223
Db 184 lrydterdw--aedqfydydvsklgrlaypp-----ealvkyrlhangvsskhsvrqh-- 234

QY 224 SAILQDFYKEYSIDITKITNYFQEVVIRKKRYT---QOELSKYFEL--KSTPSITIRKL 277
Db 235 -eiagqgiq-----tarndflqsmgfktrfslsleyrtkkaayelpkdlpeedfera 286

QY 278 YICLYLYFK 286
Db 287 rrflyqcfk 295

RESULT 12
AA749230
ID AA749230 standard; Protein; 322 AA.
XX
AC AA749230;
XX
DT 30-MAY-2000 (first entry)
XX
DE CpsII protein which has glycosyltransferase activity.
XX
KW Capsular gene cluster; serotype 1; polysaccharide biosynthesis;
KW capsular component; antigen; regulation; chain length determination;
KW complement-mediated opsonophagocytosis; serotype-specific detection;
KW antigen; vaccine; Streptococcal disease; CpsIE; CpsIF; CpsIG;
KW CpsIH; CpsII; CpsIJ; CpsIK; glycosyltransferase; CP polymerase.
XX
OS Streptococcus suis.
XX
PN WO200005378-A2.
XX
PD 03-FEB-2000.
XX
PF 19-JUL-1999; 99WO-NL00460.
XX
PR 22-JUL-1998; 98EP-0202465.
XX
PR 22-JUL-1998; 98EP-0202467.
XX
PA (DIEN-) STICHTING DIENST LANDBOUWKUNDIG ORDER20E.
XX
PI Smith HE;
XX
DR WPI; 2000-195104/17.
XX
DR N-PSDB; AA260930.
XX
PT New nucleic acid containing the capsular gene cluster of Streptococcus
PT suis, used for serotype-specific detection and to generate antigens or
PT mutants for vaccination
XX
```

```
PS Disclosure; Fig 4; 144pp; English.
XX
CC The proteins AAY68970-76 are encoded by the capsular gene cluster of
CC Streptococcus suis serotype 1. The genes in this cluster are involved
CC in polysaccharide biosynthesis of capsular components and antigens. The
CC proteins have glycosyltransferase activities (CpsIE, CpsIG, CpsII,
CC CpsIJ, CpsIK) and CP polymerase activities (CpsIH). The capsule confers
CC bacterium resistance to complement-mediated opsonophagocytosis. The
CC gene cluster is used as a source of probes and primers for
CC serotype-specific detection of S. suis and is also useful for
CC recombinant production of the proteins. The proteins are then useful
CC for producing antigens that can be used in vaccines, for controlling
CC or eradicating a Streptococcal disease, in humans or animals,
XX e.g. against S. suis in pigs.
SQ Sequence 322 AA;

Query Match 13.2%; Score 210; DB 21; Length 322;
Best Local Similarity 25.1%; Pred. No. 3.8e-11;
Matches 83; Conservative 69; Mismatches 115; Indels 64; Gaps 16;

QY 4 SIIMSYNEPLNVVRDSVESILNQTLDFEIIIVIDNPSRGDLKQFLTEYSVVDNRKIL 63
Db 6 svilpynvq-dyldkcinsliinqtytnlevilvndg-stddsekiclnymkndgrik-y 62

QY 64 LNEENIGLASSLNKAVKISKGEYIFRMDADDISYPSRFDKQIRFMEENSIDFSATLIE-- 121
Db 63 ykkinggladarnfglehatgkyiafvdsdddyevamfer----mhdnitynadiaeid 118

QY 122 --LIDQKGNLVYKQRESNKIYLT-NDIRKMLNRSILAHPTWC-----VKKKVPD 168
Db 119 fclvdeng-ytkkkkrnsfhvltreetvkeflsgsniennvwcklysrdrdiidkdkfquinn 177

QY 169 KLMGYRDLVPVE-----DYDFAIR-GALADFKIGLLNKVLQYLRNENGI 212
Db 178 rsigedllfnlevlnnvtvvvdtreyynnyvrnsslinqkfs-innidlvtrlenypf 236

QY 213 SQTNRKQVIYSAILQDFYK-----EKSYIDITKITNYFOE-----YVKKRY- 255
Db 237 kikrefshyfdakvikekvcklnkmystdclndnefipi--lesyrkeirrypfikakryl 294

QY 256 TQOELSKYFELKSTPSITIRKLYICLYLYFK 286
Db 295 srkhvltlylmkfsf-----klyvmlykfkf 320

RESULT 13
AA749203
ID AA749203 standard; Protein; 270 AA.
XX
AC AA749203;
XX
DT 22-DEC-2000 (first entry)
XX
DE Campylobacter jejuni OH4384 glycosyltransferase.
XX
KW Biosynthetic locus; biosynthesis; lipid A biosynthesis;
KW acetyltransferase; glycosyltransferase; Beta-1,4-GalNAc transferase;
KW Beta-1,3-galactosyltransferase; alpha-2,3-sialyltransferase;
KW sialic acid synthase; CMP-sialic acid synthetase; mimic; antibody;
KW immunity; immunogen; ganglioside.
XX
OS Campylobacter jejuni OH4384.
XX
PN WO200046379-A1.
XX
PD 10-AUG-2000.
XX
PF 01-FEB-2000; 2000WO-CA00086.
XX
PR 01-FEB-1999; 99US-0118213.
XX
PR 31-JAN-2000; 2000US-0495406.
```


